BINDING SERVICES
Tel +44 (0)29 2087 4949
Fax +44 (0)29 20371921
e-mail bindery@cardiff.ac.uk

THE DEVELOPMENT OF A NOVEL METHOD FOR THE CLASSIFICATION OF OSTEOARTHRITIC AND NORMAL KNEE FUNCTION

Lianne Jones, M.Eng (Hons)

PhD Thesis

2004

School of Engineering

Cardiff University

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

Advances in our understanding of human locomotion can be futile if no practical use is made of them. For the long-term benefit of patients in a clinical setting, scientists and engineers need to forge stronger links with orthopaedic surgeons to make the most use of the recent developments in motion analysis technology. With this requirement as a driving-force, an objective classification tool was developed that uses motion analysis for an application to clinical diagnostics and monitoring, namely knee osteoarthritis (OA) progression and total knee replacement (TKR) recovery.

The classification tool is based around the Dempster-Shafer (DS) theory, and as such is built upon the sound foundations of Bayesian statistics. The tool expands on the work of Safranek *et al.* (1990) and Gerig *et al.* (2000) who developed and used parts of the classification method in the areas of vision and medical image analysis respectively. Using the data collected during a clinical knee trial, this novel approach enables the objective classification of subjects into an OA or normal group. Each piece of data is transformed into a set of belief values: a level of belief that a subject has OA knee function, a level of belief that a subject has NL knee function and an associated level of uncertainty. The belief values are then represented on a simplex plot, which enables the final classification of a subject, and the level of benefit achieved by TKR surgery to be visualised. The DS method can be used as a fully or partially automated tool. The input variables and control parameters, which are an intrinsic part of the tool, can be chosen by an expert or an optimisation approach.

Using a leave-one-out (LOO) approach, the tool was able to classify new subjects with an accuracy of 97.62%. This compares with the 63.89% and 95.24% LOO accuracies of two well-established methods – the Artificial Neural Network and the Linear Discriminant Analysis classifiers respectively. The tool also provides an objective indication of the variables that are the most influential in distinguishing OA and NL knee function. In this case, the variables identified by the tool as important are often cited as clinically relevant variables, which enhances the appeal of the tool to the clinical community and allows for more effective comparison with clinical approaches to diagnosis. Using Simulated Annealing to select the control parameters reduced the LOO accuracy to 95.24%. Automated feature selection using a Genetic Algorithm and Sequential Forward Selection increased the LOO accuracy to 100%. However, further work is required to improve the effect of this process on the overall level of uncertainty in the classification.

Initial studies have demonstrated a practical and visual approach that can discriminate between the characteristics of NL and OA knee function with a high level of accuracy. Further development will enable the tool to assist orthopaedic surgeons and therapists in making clinical decisions, and thus promote increased confidence in a patient's medical care.

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed.....Lianne Jones.....(candidate)
Date.....21/6/04.....

STATEMENT 1

This thesis is the result of my own investigation, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed.....Lianne Jones.....(candidate)
Date.....21/6/04.....

STATEMENT 2

I hereby give consent for my thesis, if, accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed.....Lianne Jones.....(candidate)
Date.....21/6/04.....

ACKNOWLEDGEMENTS

I would like to express my thanks to my supervisor Dr. Cathy Holt for all her input over the last few years. Her help, encouragement, vision, patience and energy have been immense. I have learnt so much from working with her, and I am extremely grateful for her constant belief in me and my research.

Secondly, I would like to thank Dr. Malcolm Beynon for his enthusiasm and help with all the Dempster-Shafer work. I have greatly appreciated his input into this thesis. His advice and guidance have always sparked off many new avenues of thought and research.

Thirdly, I am hugely grateful to all colleagues, friends and patients who took part in the clinical trial and to the orthopaedic surgeons Stuart Roy and Richard Evans for sharing their time to give a clinical perspective on my results and for ploughing through many hours of clinical gait analysis videos.

Finally, I would like to thank friends and family for their support throughout this work. I cannot finish without giving a special mention to my parents, Ian and Denise Jones, and to two of my best friends Christopher Cruise and Ruth Parkes. I am certain that without their constant support, encouragement, thoughtfulness, patience and ability to make me laugh, I would never have completed this thesis.

CONTENTS

LIST OF FIGURES

CHAPTER 4

CHAPTER 5

CHAPTER 6

CHAPTER 7

APPENDIX B

LIST OF TABLES

CHAPTER 5

CHAPTER 6

CHAPTER 7

APPENDIX B

APPENDIX C

ABBREVIATIONS

AAR	Abduction-adduction rotation
ALCS	Anatomical local coordinate system
AAR	Abduction-adduction rotation
ACL	Anterior cruciate ligament
ANN	Artificial Neural Network
AP	Anterior-posterior
APF	Anterior-posterior ground reaction force
BFOD	Binary frame of discernment
BMI	Body mass index
BOA	Blinded Observational Analysis
BOE	Body of evidence
BOE_c	Combined body of evidence
CA	Cluster analysis
CMA	Clinical motion analysis
CPU	Computer processing unit
CS	Coordinate system
DS	Dempster-Shafer
DST	Dempster-Shafer theory
EMG	Electromyography
FA	Fourier Analysis
FER	Flexion-extension rotation
FOD	Frame of discernment
FTKR	Fixed bearing total knee replacement
GA	Genetic Algorithm
GAANN	Genetic Algorithm Artificial Neural Network

GCS	Global coordinate system
GRF	Ground reaction force
IER	Internal-external rotation
JCS	Joint coordinate system
KOS	Knee outcome survey
LCS	Local coordinate system
LDA	Linear Discriminant Analysis
ML	Medial-lateral
MLCS	Marker cluster local coordinate system
MSE	Mean squared error
NL	Normal
OA	Osteoarthritis/Osteoarthritic
OB	Objective Function
PC	Principal Component
PCA	Principal Component Analysis
PCL	Posterior cruciate ligament
PLCS	Pointer local coordinate system
P1	Patient 1
P2	Patient 2
P3	Patient 3
P4	Patient 4
P5	Patient 5
P6	Patient 6
P7	Patient 7
P8	Patient 8
P9	Patient 9
RTKR	Rotating platform total knee replacement

RW	Raised walkway
SBS	Sequential Backward Selection
SFS	Sequential Forward Selection
SLDA	Stepwise Linear Discriminant Analysis
SSM	Sequential Selection Methods
TKR	Total knee replacement
V	Vertical
VC	Video camera
VF	Vertical ground reaction force

NOTATION

CHAPTER 1

a	Discriminant function constant
a_{ji}	PC loading
a_0	Constant Fourier coefficient
a_k	Constant Fourier coefficient
b_i	Discriminant function coefficient
b_k	Constant Fourier coefficient
$A(\omega)$	Fourier coefficient
$B(\omega)$	Fourier coefficient
D	Discriminant score
$g(t)$	Periodic function
p	Number of input variables
$x(t)$	Non-periodic function
$X(\omega)$	Fourier transform of $x(t)$
v_i	Input variable
ω	Fundamental frequency

CHAPTER 2

$\mathbf{A}$	Singular value decomposition vector
$\mathbf{B}$	Singular value decomposition vector
c_{kl}	Correlation coefficient between variables z_k and z_l
$\mathbf{C}$	Correlation matrix
$\mathbf{d}$	Translation vector
$\mathbf{D}$	Singular value decomposition matrix
$[D_F]$	Force calibration matrix

e_j	Eigenvector
$\mathbf{E}$	Matrix of eigenvectors
f_i	Vector of coordinates of a body in the CS, CS1
$\bar{f}$	Mean of the coordinates of a body in the CS, CS1
F_{ML}	Medial-lateral GRF
F_{AP}	Anterior-posterior GRF
F_V	Vertical GRF
F_x	Force component in the force platform CS
F_y	Force component in the force platform CS
F_z	Force component in the force platform CS
g_i	Vector of coordinates of a body in the CS, CS2
$\bar{g}$	Mean of the coordinates of a body in the CS, CS2
h	Height
l_{ij}	Correlation coefficient between variable i and PC j
$\mathbf{L}$	Component loadings matrix
m	Number of PCs retained
M_x	Moment component in the force platform CS
M_y	Moment component in the force platform CS
M_z	Moment component in the force platform CS
M_i	Marker cluster marker
n	Number of subjects
$\mathbf{n}_i$	Unit vector used to establish femoral ALCS
$\mathbf{N}_i$	Vector used to establish the femoral ALCS
$origin_{a,m}$	MLCS coordinates of the ALCS origin
$origin_{m,g}$	GCS coordinates of the PLCS origin
$origin_{p,g}$	GCS coordinates of the MLCS origin

p	Number of input variables
$\mathbf{P}$	Singular value decomposition matrix
P_i	Pointer cluster marker
P_m	The MLCS coordinates of the pointer's point
P_p	The PLCS coordinates of the pointer's point
$\mathbf{Q}$	Singular value decomposition matrix
$\mathbf{R}$	Rotation matrix
$\mathbf{S}$	Singular value decomposition matrix
t_m	Total variance accounted for by m PCs
$\mathbf{T}_{pg}$	Transformation matrix relating orientation of the PLCS in the GCS
$\mathbf{T}_{mg}$	Transformation matrix relating orientation of the MLCS in the GCS
$\mathbf{T}_{gm}$	Inverse of the matrix $\mathbf{T}_{mg}$
$\mathbf{T}_{ma}$	Transformation matrix relating orientation of the M LCS in the ALCS
$\mathbf{T}_{am}$	Inverse of the matrix $\mathbf{T}_{ma}$
$\mathbf{u}_i$	Unit vector used to establish the PLCS
$\mathbf{U}_i$	Vector used to establish the PLCS
$\mathbf{v}_1$	Unit vector used to establish the MLCS
$\mathbf{V}_1$	Vector used to establish the MLCS
w	Weight
$\mathbf{w}_i$	Unit vector used to establish the tibial ALCS
$\mathbf{W}_i$	Vector used to establish the tibial ALCS
z_{ij}	Standardised variables
$\mathbf{Z}$	Matrix of standardised variables
Γ	Matrix of singular values
φ_i	Singular values
θ_{FE}	Flexion-extension angle
θ_{AA}	Abduction-adduction angle

θ_{IE}	Internal-external rotation angle
λ_j	Eigenvalue
$\mathbf{\Lambda}$	Diagonal matrix of eigenvalues
σ_j	Sample standard deviation for y_{ji}
$\mathbf{\Omega}$	Matrix of PC scores

CHAPTER 3

A	DS control parameter
B	DS control parameter
bpa	Basic probability assignment
$cf(v)$	Confidence factor
i_t	SA control parameter, iterations per temperature
k	DS control parameter
k_c	DS control parameter k
k_s	DS control parameter k
$m(.)$	Probability mass function (*bpa*)
$m_c(.)$	Combined probability mass function (*bpa*)
n_i	Number of input variables
{NL}	The hypothesis "the subject has NL knee function"
{OA}	The hypothesis "the subject has OA knee function"
$p(v_j\|w_i)$	Conditional probability
$p(w_j)$	Prior probability
$p(w_j\|v_i)$	Posterior probability
r_t	SA control parameter, temperature reduction factor
t_f	SA control parameter, final temperature
t_s	SA control parameter, starting temperature
t	Temperature

v	Input variable
$\{x\}$	the hypothesis "the object is at its predicted position"
$\{\neg x\}$	the hypothesis "the object is not at its predicted position"
Δf	Difference between current OB and new OB
θ	DS control parameter
Θ	Frame of discernment
Θ_L	Lower uncertainty boundary
Θ_U	Upper uncertainty boundary
κ	Normalisation factor
$\{\varphi_i\}$	Hypothesis
$\{\phi_i\}$	Elementary hypothesis
$\varnothing$	The null hypothesis or empty set

CHAPTER 4

A	DS control parameter
B	DS control parameter
$cf(v)$	Confidence factor
i_t	SA control parameter – iterations per temperature
k	DS control parameter
k_c	DS control parameter k, calculated using correlation coefficient method
k_s	DS control parameter k, calculated using standard deviation method
$m_c(.)$	Combined probability mass function (*bpa*)
{NL}	The hypothesis "the subject has NL knee function"
{OA}	The hypothesis "the subject has OA knee function"
r_t	SA control parameter –temperature reduction factor
t_s	SA control parameter – starting temperature

t_f	SA control parameter – final temperature
v_i	Input variable
Δf	Difference between current OB and new OB
θ	DS control parameter
Θ	Frame of discernment
Θ_L	Lower uncertainty limit
Θ_U	Upper uncertainty limit

CHAPTER 5

A	DS control parameter
B	DS control parameter
d_0	Discriminant function constant term
d_i	Discriminant function coefficient
D	Discriminant score
$f(w_j)$	Fitness value for chromosome w_j
$F(w_j)$	Probability of chromosome w_j being selected
m	Population size
$m_c(.)$	Combined probability mass function
n_{sel}	Number of chromosomes selected
{NL}	The hypothesis "the subject's knee function is normal"
{OA}	The hypothesis "the subject's knee function is osteoarthritic"
OB	Euclidean distance objective function
OB_{FS}	Feature selection objective function
OB_{IN}	In-sample Euclidean distance objective function
p	Number of input variables
p_c	Crossover probability
p_m	Mutation probability

p_{sel}	Random number
q_j	Upper limit for generating random number p_{sel}
r_j	Ranking of the chromosome w_j
s	Selected pressure
SS_{NL}	NL group sum of squares matrix
SS_{OA}	OA group sum of squares matrix
SS_t	Total sum of squares matrix
SS_w	Within-group sum of squares matrix
$SSCP_t$	Total sums of squares and cross products matrix
$SSCP_w$	Within-group sums of squares and cross products matrix
v_i	Input variable
V	Set of input variables
w_j	Chromosome
α	Average in-sample classification error
Λ	Wilks' lambda statistic
Θ	Frame of discernment/Uncertainty
Θ_L	Lower uncertainty limit
Θ_U	Upper uncertainty limit

CHAPTER 6

A	DS control parameter
B	DS control parameter
k	DS control parameter
$m_c(.)$	Combined probability mass function
{NL}	The hypothesis "the subject's knee function is normal"
{OA}	The hypothesis "the subject's knee function is osteoarthritic"
p	Significance level

r	Correlation coefficient
θ	DS control paramter
Θ	Frame of discernment
Θ_L	Lower uncertainty limit
Θ_U	Upper uncertainty limit

CHAPTER 7

a_j	Weighted sum of all inputs to hidden neuron
b_j	Weighted sum of all inputs to output neuron
c_{j0}	Constant classification function term
c_{ji}	Classification function coefficient
$\mathbf{C}_j$	Matrix of classification coefficients
d_0	Constant term in discriminant function
d_i	Discriminant function coefficient
D	Discriminant score
E	Error function
$f(.)$	Activation function
H_i	Hidden neuron
I_i	Input neuron
$m(.)$	Probability mass function
$\mathbf{M}_j$	Matrix of means
n_i	Number of input variables
n_j	Number of hidden neurons
n_k	Number of output neurons
n_p	Number of patterns in the training set
O_k	Output neuron
p	Significance level

OB_{rank}	Ranking objective function
RS_{ki}	Relative strength of the relationship between input variable v_i and the output unit O_k
S_j	Classification score
t_k	Target output
U_{jk}	Connection weight connecting hidden and output neurons
$\mathbf{VC}_w$	Within-groups variance-covariance matrix
W_{ji}	Connection weight connecting input and hidden neurons
v_i	Input variable
y_k	Network output
z_j	Output from hidden neuron
δ_j	Error for output neuron
δ_k	Error for hidden neuron
η	Learning rate
μ	Momentum constant
Θ	Frame of discernment
Θ_L	Lower uncertainty limit
Θ_U	Upper uncertainty limit

CHAPTER 8

A	DS control parameter
B	DS control parameter
k_c	DS control parameter k calculated using correlation coefficient method
k_s	DS control parameter k calculated using standard deviation method
Θ_L	Lower uncertainty boundary
Θ_U	Upper uncertainty boundary

CHAPTER 9

A	DS control parameter
B	DS control parameter
Θ_L	Lower uncertainty boundary
Θ_U	Upper uncertainty boundary

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 INTRODUCTION

Surgeons and physiotherapists use a range of simple observations and physical examinations to decide on the extent of a patient's illness and to proceed with a diagnosis and plan of surgery or therapeutic treatment. Their decisions are based on experience of patients with similar symptoms and their expertise and training.

However, following the emergence and development of new Clinical Motion Analysis (CMA) techniques over the last 15 years, the global scientific community has begun to realise that the field of biomechanics and human motion analysis has much to contribute to the development of new and emerging medical diagnostic techniques.

The development of CMA techniques as a non-invasive aid to medical assessment has important implications in the area of patient diagnosis. With close partnership between orthopaedic surgeons, scientists and engineers, the tools available for patient assessment and diagnosis are set to improve and patients will benefit from improved accuracy and efficiency of their diagnosis, promoting increased confidence in their medical care.

This thesis stems from such a desire to help with and improve the process of patient diagnosis that will be beneficial to the health of the general population in the long-term. As such the purpose of this thesis is to provide the foundations for developing a sophisticated diagnostic tool to aid orthopaedic surgeons and therapists when making clinical decisions based upon CMA techniques.

Patients with movement disorders are regularly referred to CMA laboratories. During these brief assessments, a wealth of biomechanical data relating to the functional abnormality is collected, e.g., range of movement etc. However, it is the experience of the author and other colleagues that it is extremely difficult to objectively analyse and

gain conclusions from such a wealth of data. This is reinforced by the work of Benedetti *et al.* (1998) who commented that "it is often not easy for the clinician to examine so much data without a systematic and rigorous approach" (pp.204).

As a result, the aim of this research is to develop an objective diagnostic tool that would be capable of producing an automated diagnosis from the CMA data. The objective diagnostic tool would be developed in conjunction with orthopaedic surgeons and thus make a decision which is in-line with clinical opinion. Such an objective tool would have significant clinical value as it could provide useful information on the effectiveness of pre-operative disease progression, surgery and post-operative therapy and recovery assessment. With an improvement to the clinical assessment process for common diseases, surgery to relieve the painful and functionally disabling symptoms could be more effectively tailored to suit patients. The clinical tool could provide a powerful prediction of the extent to which a patient presenting a distinct set of pre-operative symptoms would respond to various treatment options.

The remainder of this thesis is concerned with the development of an objective classification tool using motion analysis for a proposed application to clinical diagnostics and monitoring. A brief introduction is provided to the historical development of motion analysis in general, followed by a description of the proposed application of the objective tool. A detailed review of the previous techniques and research carried out in the area of objective gait analysis is given and the new objective tool is then introduced with a number of key objectives.

1.1.1 A brief history of gait data analysis

The analysis of human movement dates back many centuries. Indeed, even as early as the 4th century BC philosophers such as Aristotle were providing qualitative descriptions of human locomotion. However, it was many centuries later until scientists began to quantitatively measure human movement.

In the 17th century, scientists including Galileo, Newton and Borelli began to apply the principles of motion to human locomotion (Whittle, 1996). In the early 19th

century, the Weber brothers produced quantitative measurements of temporal-distance parameters (Whittle, 1996; Andriacchi and Alexander, 2000). Towards the end of the 19th century and the beginning of the 20th century, Marey, Muybridge and Braune and Fischer developed pioneering techniques to quantitatively measure human motion (Whittle, 1996; Andriacchi and Alexander, 2000). Marey developed a number of devices to record movement including the photographic rifle and an early version of a force plate. Muybridge produced sequences of photographic images of animals and humans in motion (Whittle, 1996; Andriacchi and Alexander, 2000). Braune and Fischer developed a photographic technique that could track movement in 3-D and measurements to describe movement in terms of kinematics and kinetics (Andriacchi and Alexander, 2000; Whittle, 1996). However, the calculations that they made were extremely complicated and time consuming.

Since these early pioneering works the science of gait analysis has progressed considerably. With the introduction and advancement of computer technology the use of instrumented gait analysis in a research setting has become widespread. The modern gait analysis laboratory typically houses a set of digital and video cameras to capture dynamic movement; digital force platforms to measure kinetic data; and EMG equipment to record muscle activity. It is now possible to obtain and process highly accurate measurements of sophisticated movement in the space of a few hours. In contrast to the pioneers of motion analysis, although there are questions that relate to the degree of accuracy of the motion analysis methods, the main challenge facing the biomechanical community today is no longer how to produce data that quantifies human movement; this is now commonplace. Rather, one of the greatest challenges is how to use this information so that it can be of clinical benefit (Whittle, 1996).

1.1.2 Context and aims of work

This work stems from an on-going clinical trial conducted in the Cardiff University motion analysis laboratory. The trial was established to examine the knee function of patients before and after surgery to treat their osteoarthritis (OA).

Knee OA is a common degenerative disease that affects the cartilaginous surfaces of the joint. The cartilage roughens and becomes thinner and, as a result, the surrounding

bone thickens, osteophytes form and the synovium swells (Arthritis Research Campaign, 1998). Additionally, the surrounding soft tissues contract and the muscles atrophy. This deformity limits the movement of the joint and often causes a patient severe pain. In cases of severe knee OA, total knee replacement (TKR) surgery is performed to remove the diseased surfaces of the joint and replace them with prosthetic components. In most cases, TKR surgery reduces knee pain and restores a degree of normal (NL) knee function (Andriacchi, 1993; Myles *et al.*, 2002).

The knee clinical trial in Cardiff was specifically put in place to characterise the differences between two different Depuy TKR designs: the P.F.C. Sigma (fixed-bearing knee, (FTKR)) and the DePuy P.F.C. Sigma R.P. (rotating platform knee, (RTKR)) based on the measurements taken during the laboratory sessions. To date, comparison of the two TKR designs at different stages of post-operative recovery has been achieved by examining the transverse plane range of motion of the tibiofemoral joint and subsequently comparing them to the mean of a group of NL knees (Holt *et al.*, 2002, Roy *et al.*, 2002), as shown in Figure 1.1. However, there are problems associated with this simplistic approach. These main concerns are now discussed.

During a session in the Cardiff motion analysis laboratory, anthropometrical measurements, kinetic and kinematic waveforms, temporal-distance parameters and patient related information are collected. A summary of the variables recorded is given in Figure 1.2. The vast amount of data obtained from such an analysis is common practice as highlighted by the work of Benedetti *et al.* (1998) who collected 124 parameters to describe each subject's gait.

Despite collecting such a mass of data during these sessions, analysis so far has been restricted to one single variable: range of motion in the transverse plane. However, among all the variables collected during a typical gait analysis, no individual variable is capable of providing a complete description of a subject's gait (Jacobs *et al.*, 1972). Furthermore, use of one variable in isolation to the rest has meant that a vast amount of potentially important information lies redundant. Rather than identify the changes in individual variables, it would be of value to see how the alteration of multiple variables has combined to produce an overall transformation in a subject's gait pattern.

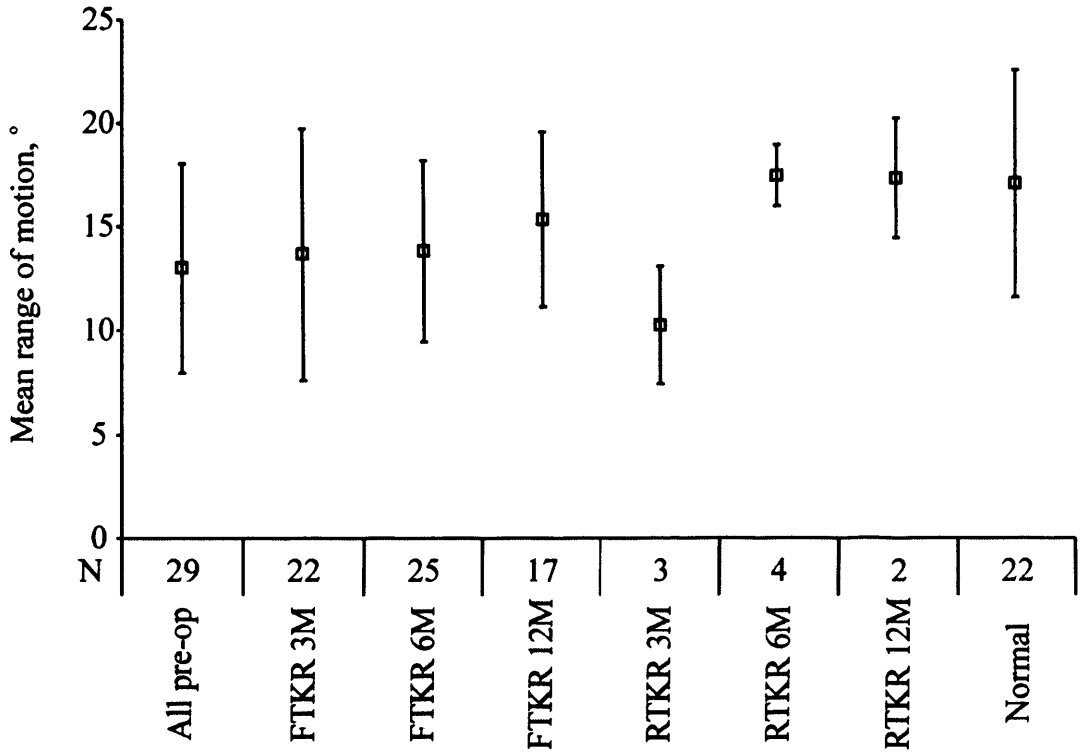


Figure 1.1 Mean external-internal tibial range of motion (± 1 standard deviation) for pre-op, FTKR and RTKR post-op and normal subjects. M = Number of months post-op, RTKR = Rotating platform TKR, FTKR = Fixed bearing TKR, N = sample size (Reproduced from Holt *et al.*, 2002).

A high proportion of the data recorded during the gait analysis sessions exists in the form of temporal waveforms. As mentioned above, these temporal waveforms are currently parameterised (Holt *et al.*, 2002; Roy *et al.*, 2002); a practice that is prevalent in many other gait analysis studies (as an example see Chao *et al.*, 1980 and Benedetti *et al.*, 1998). A danger associated with this practice is highlighted in Figure 1.3, which depicts the internal-external rotation waveforms of two NL subjects. Although the ranges of motion of the two subjects are very similar (9.9° and 9.8° respectively) the shape of the waveforms is very different. Consequently, there is a danger that valuable temporal information is discarded. Furthermore, gait parameters defined using NL temporal waveforms are not always easily identifiable in pathological waveforms (Chau, 2001a; Deluzio *et al.*, 1997; Chao *et al.*, 1983).

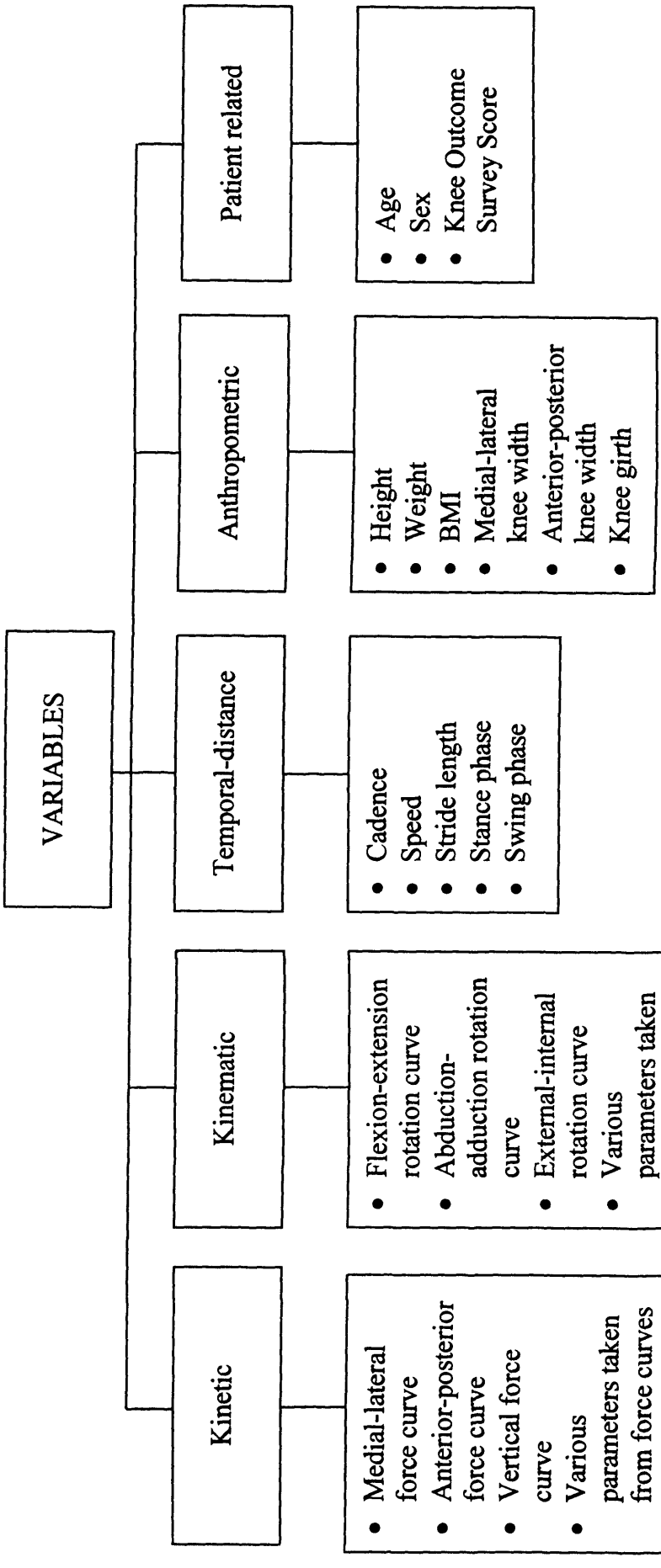


Figure 1.2 Variables collected during gait analysis session

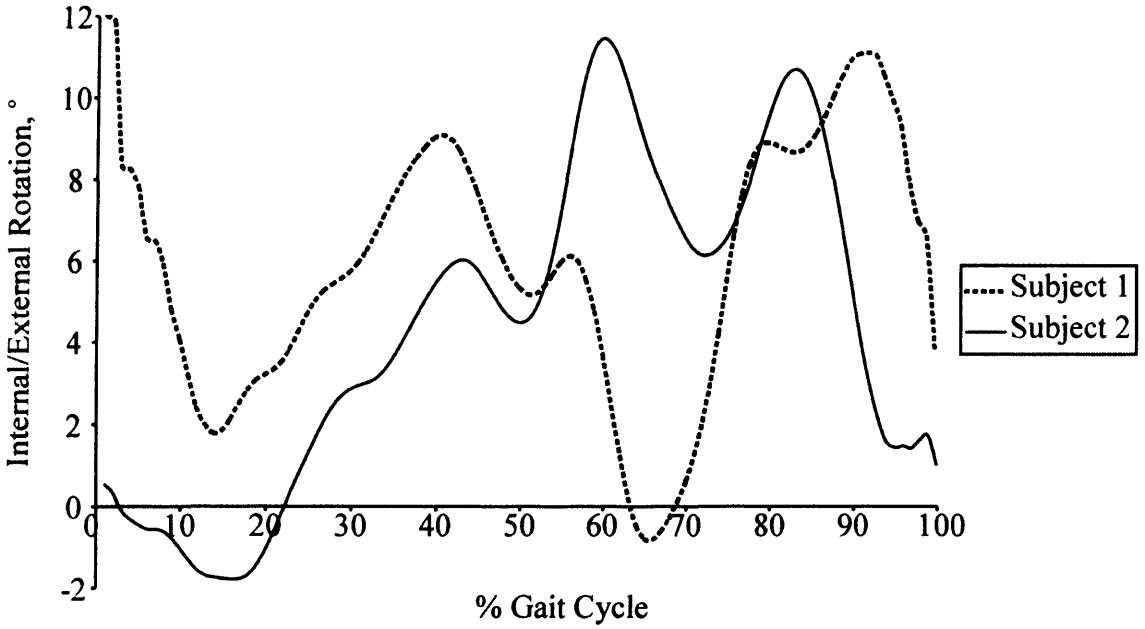


Figure 1.3 Internal-external rotation waveforms for two different subjects. The range of values for the two waveforms is the same but the overall shape is different.

In response to the inadequacies of the present analysis technique, a need has been identified to develop a new analysis method to characterise the differences between NL, OA (pre-operative TKR) and post-operative TKR subjects.

Such a characterisation or classification method would be useful in the determination of pathology and improvement due to surgical and therapeutic intervention. Practical applications of a simple characterisation method would include monitoring of joint degeneration and diagnostics; prediction of outcome for surgical intervention; post-operative monitoring and functional analysis of joint prosthesis design.

The next section of this chapter seeks to present a comprehensive description of the current methods used to analyse gait data and in so doing identify their capabilities, limitations and assess their applicability to the classification of OA, NL and TKR knee function.

1.2 LITERATURE REVIEW

Chau (2001a; 2001b) presents an extensive review of the emergent techniques applied to the task of gait data analysis, namely Multivariate Statistics, Fuzzy methods,

Wavelet Analysis, Fractal Analysis and Artificial Neural Networks (ANN). Other methods reported in the literature include Fourier Analysis (FA), Linear Discriminant Analysis (LDA) and Cluster Analysis (CA). Each method that is relevant to the current problem is now considered in turn. In each case, a brief summary of the method is followed by a review of the applications of the method. Finally, in section 1.2.5 a discussion is given on the capabilities and limitations of each method within the context of the current problem.

1.2.1 Fourier Analysis (FA)

The Fourier transform is a frequency-domain representation of a function. The representation in the frequency domain contains exactly the same information as that of the original function; they differ only in the matter of presentation. Periodic functions, $g(t)$, can be expressed as an infinite sum of sines and cosines at discrete harmonic frequencies of the fundamental period of the signal:

$$g(t) = a_0 + \sum_{k=1}^{\infty} \left(a_k \cos(k\omega t) + b_k \sin(k\omega t)\right) \tag{1.1}$$

where a_0 is a constant, a_k and b_k are Fourier coefficients, and ω the fundamental frequency (Stroud, 1984).

Non-periodic functions cannot be broken down into discrete frequency components. However, such functions can still be represented in the frequency domain. In this case, the Fourier series becomes a Fourier integral and the Fourier coefficients are continuous functions of frequency, called Fourier transforms:

$$x(t) = \int_{-\infty}^{\infty} X(\omega) e^{-i\omega t} dt \tag{1.2}$$

where $X(\omega) = A(\omega) - iB(\omega)$ is the Fourier transform of $x(t)$ (Brigham, 1988). Thus, waveform data can be expressed in terms of discrete parameters.

Jacobs *et al.* (1972) first considered the value of FA in the analysis of gait data. Choosing the ground reaction force (GRF) as their working parameter, a FA representation of the vertical GRF waveform was obtained for 25 NL subjects and 239 pathological hip patients. The GRF waveforms and the condensed harmonic representations were examined visually and an attempt was then made to relate these to the diagnosis of the patient. The harmonic spectra of the NL waves were similar over the first six harmonics but varied considerably beyond this. The harmonic spectra and GRF waveforms themselves varied largely for the patients within the pathological sample; hence generalising about pathological gait was more complex a task. Eight different waveform types were identified by visual examination of the GRF waveforms. Although this was based on subjective opinion, some correlation was found between their features, the harmonic spectra and the diagnostic state of the hip. This use of FA highlights its potential in providing succinct representations of continuous waveform data to enable easier comparison of waveform data.

Chao *et al.* (1983) established a database of temporal-distance parameters, knee joint motion and GRF waveforms for 148 normal adults during level walking. FA was also applied to the knee joint motion and GRF waveforms and its use again highlights its potential in providing representations of the entire gait cycle data.

Following on from the work of Chao *et al.* (1983), Schneider and Chao (1983) performed a study to examine the potential of FA in distinguishing NL and pathological gait. FA was performed on the GRF waveforms of 26 NL subjects and 10 TKR patients. A method was developed to determine the essential number of harmonics required to reconstruct the original GRF waveforms. It was found that the first two to four harmonics were the dominating coefficients in describing each of the GRF patterns. Statistical analysis revealed a difference in the corresponding harmonic coefficients of the NL and OA patterns. These findings led Schneider and Chao (1983) to suggest that Fourier coefficients could be used as key parameters to differentiate the two groups. Indeed, in later classification studies many researchers used Fourier transformed variables as inputs to ANN (Holzreiter and Köhle, 1993; Köhle and Merkl, 1996; Barton and Lees, 1997) and to Principal Component models (Borzelli *et al.*, 1999).

FA can produce discrete representations of continuous gait waveforms whilst retaining temporal information (Jacobs *et al.*, 1972; Chao *et al.*, 1983; Schneider and Chao, 1983). The most successful application of FA has been as a pre-processing method to other analysis techniques (Holzreiter and Köhle, 1993; Köhle and Merkl, 1996; Barton and Lees, 1997; Borzelli *et al.*, 1999).

1.2.2 Multivariate Statistics

Principal Component Analysis (PCA), Factor Analysis, Linear Discriminant Analysis (LDA) and Cluster Analysis (CA) are all well-established, Multivariate Statistical methods. Together, they are the most widely understood and applied in the analysis of gait data (Chau, 2001a; 2001b). All of the reviewed studies utilising Factor Analysis are concerned with the identification of EMG patterns and consequently are not considered here (see Davis and Vaughan, 1993; Olree and Vaughan, 1995; Merkle *et al.*, 1998)

1.2.2.1 Principal Component Analysis (PCA)

PCA is a Multivariate Statistical method that aims to obtain a succinct representation of a high dimensional data set of interrelated variables. PCA transforms the data set into a new set of uncorrelated variables, or components, retaining the variation present in the original data set. A more comprehensive description of PCA is given in Chapter 2 (section 2.5.1), but a concise summary is given here.

The Principal Components (PCs) are linear combinations of the original variables v_i (i = 1:p). The jth PC, P_j is given by,

$$P_j = \sum_{i=1}^{p} a_{ji} v_i \qquad (1.3)$$

where a_{ji} are the component loadings, which are indicative of the amount of variation in variable v_i described by the PC, P_j. Although PCA produces the same number of PCs as there are original variables it is anticipated that the variances of the majority of

PCs will be negligible and that the original variables can be described by a smaller number of PCs.

The first application of PCA in the analysis of gait data was by Shiavi and Griffin (1981) whilst investigating the EMG and foot-contact patterns of NL subjects. The gait cycle was divided into 16 time segments and a 16-element vector was constructed for both the EMG and foot-contact patterns. Each element value depicts the proportion of time that either a muscle is active or a foot makes contact. Shiavi and Griffin (1981) used PCA to reduce the dimensionality of the data, thus enabling the clustering of similar patterns. This early work revealed the potential of PCA for gait data reduction and for use as a pre-processing step to CA.

Yamamoto *et al.* (1983) used PCA to quantitatively evaluate the recovery of patients with hip disease. Joint angular displacement (sagittal plane only), GRFs, point of force application and temporal-distance factors were measured for 115 NL subjects and 211 patients undergoing varied surgical procedures for hip disease. Using four of these measured variables and six derived representations of them, ten items were chosen as inputs to the PC model. Three out of the ten PCs were retained, although no comment is made as to how much of the total variation these three components account for. Yamamoto *et al.* (1983) went further than Shiavi and Griffin (1981) and interpreted their PCs, thus maintaining the clinical relevance of their method. Through examination of the component loadings, the first two PCs were labelled 'total gait ability' and 'activity and symmetry' respectively. The third PC was discarded as irrelevant because only three out of the ten components loaded highly against it. PCA was able to sufficiently reduce the data to enable a visual representation of results. The first two PCs were plotted against each other to form a 'gait evaluation plane' as shown in Figure 1.4. In projecting the data for each NL and patient (pre and post operative) subject onto this component plane, Yamamoto *et al.* (1983) were able to represent visually the variation in gait due to age and pathology, and improvements in gait due to surgical intervention (see Figure 1.5). All patients and NL subjects could be evaluated on the same scale.

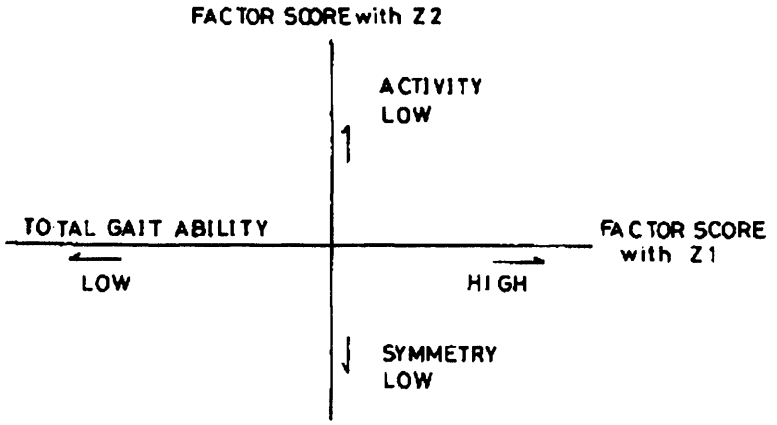


Figure 1.4 Gait evaluation plane formed by two PCs, Z_1 and Z_2 (Taken from Yamamoto *et al.*, 1983).

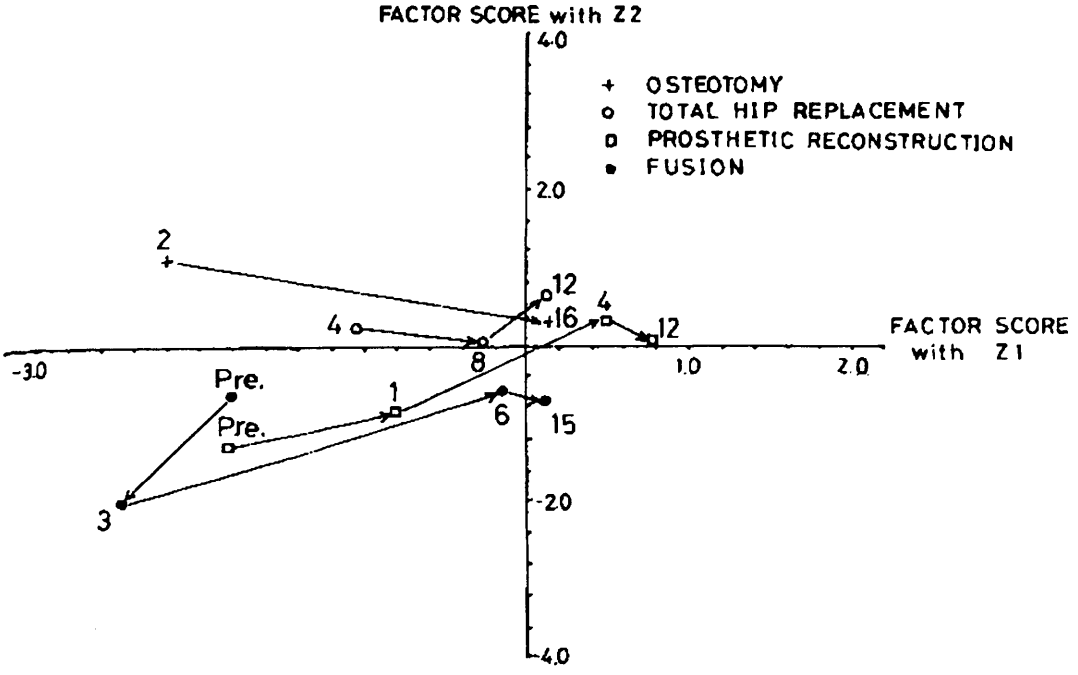


Figure 1.5 Recovery of the factor score by operation in the 'gait evaluation plane'. Numerals indicate months after operation (Taken from Yamamoto *et al.*, 1983).

Unlike previous studies, Deluzio *et al.* (1997, 1999) did not use temporal-distance parameters extracted from the gait waveforms as inputs to their PC models, but the entire gait waveform themselves; a time normalised curve sampled at each 1% from 0 to 100% of the gait cycle). Consequently, temporal information was retained. PC models were built from a dataset comprising elderly NL volunteers. PC models were developed for the three components of the knee bone on bone forces, the three net reaction moments and the two components of the relative knee angles. The number of PCs used in each model was determined using cross validation. Once the PC models had been established using the NL cohort, the gait data from patients were introduced to the PC models, converted into a set of PC scores and compared to 95% confidence limits of the NL group. Using this procedure, Deluzio *et al.* (1997, 1999) were able to detect deviations from NL and identify the portion of the gait cycle responsible for such differences.

Sadeghi *et al.* (1997) used PCA to identify the peak muscle power and energies developed by the lower limbs and to determine the relationship between these parameters and the functions of support and propulsion during NL gait. The temporal gait parameters, peak powers and mechanical energies of nineteen, right-leg dominant, NL males were analysed using PCA. PCs with eigenvalues less than one were discarded, and of the remaining PCs, those containing at least sixty percent of the information were retained. This resulted in four PCs being kept for the right limb, and four for the left. Sadeghi *et al.* (1997) did not attempt to interpret or label the PCs. Alternatively the PC loadings were studied to determine the most important gait parameters, i.e. those parameters contributing most significance to the overall variance of the data; in this instance PC loadings greater than 0.6. Ten important gait parameters were identified for the right leg relating to propulsive functions, and fifteen for the left relating to supportive functions.

Borzelli *et al.* (1999) applied PCA as a data reduction technique to a set of GRFs recorded during a number of sit-to-stand trials. Five trials were performed for 24 young and 58 elderly healthy subjects. During each trial, three force and three moment waveforms were recorded. Each signal was filtered, Fast-Fourier transformed and normalised before performing PCA. Generally, the first two PCs described over 80% of the total variation present in the original data, which emphasizes the ability of

PCA to deal with the dimensionality problem evident in gait analysis. Borzelli *et al.* (1999) examined the nature of the correlation coefficient between inter and intra-subject PCs. A high correlation for the first PC (PC1) values obtained from different trials and different subjects suggest that the whole population accomplish the motor task by adjusting the GRF in such a way to produce the same PC1. As a result, the group related PC1 to intrinsic aspects of the motor task. Low levels of correlation for the second PC (PC2) suggest that it is related to inter-subject features used to achieve the motor task. Borzelli *et al.* (1999) did not attempt any pattern classification, but suggest that the individual features highlighted by PC2 could be used in such a task.

Sadeghi *et al.* (2000) used PCA to characterise the main functional actions of the muscles at the hip during able-bodied gait and to identify any non-symmetrical behaviour between the right and left lower limbs. Twenty young, healthy males participated in the study. Following the work of Deluzio *et al.* (1997, 1999), Sadeghi *et al.* (2000) used the entire gait waveform as input to their PC model; the muscle powers at the hip (sagittal plane) at 1% increments of the gait cycle (101 data samples). PCs with eigenvalues less than one were discarded and finally the first four PCs, accounting for over 70% of the variation in the original data set, were retained. When interpreting the PCs, the parameters that had the highest correlation with each PC were used; in this instance those with component loadings greater than 0.6. The PCs represented different actions taken by the muscles during the gait cycle namely control balance, propulsion, between limb coordination and limb preparation.

Schutte *et al.* (2000) highlighted a further potential of PCA, using it to derive a gait index to quantify deviations from NL gait. The use of such a multivariate technique enabled them to examine the collective effect of surgical intervention on all gait features, instead of simply the change in individual features themselves. The sixteen parameters included in the PCA analysis were subjectively chosen based on expert clinical opinion. PC expressions were derived using a set of NL subjects. Subsequently the data belonging to the pathological sample was transformed into PC values using the PC expressions derived from the NL group. The sum of the square of the distance of the 16 PC values was interpreted as the deviation of the subject's gait from NL, known as the normalcy index. The normalcy index was used to distinguish between NL subjects and patients, and between patients with different severities of

gait abnormalities. Schutte *et al.* (2000) found that the mean index values were found to increase as gait abnormality increased. This study further highlights the potential of PCA to classify gait differences and to identify changes due to surgical intervention.

Building on the work of Deluzio *et al.* (1997, 1999), Astephen *et al.* (2002a, 2000b) developed a PC method that concurrently considered constant gait parameters and time-varying gait waveforms. The full gait cycle data of nine waveforms (three components of knee joint angles, moments and forces) and five constants (static hip knee ankle angle, standing knee flexion angle, medial and lateral condyle joint spaces and BMI) were simultaneously included in a PCA. The technique was able to discriminate between OA and NL gait patterns. Using a stepwise linear discrimination procedure, the features most able to discriminate between the two groups were identified. These features were interpreted through examination of the contribution of each of the input variables to each feature. This study revealed that the loading response phase of the gait cycle is an important factor in isolating OA.

In all the reported studies, PCA was able to significantly reduce the dimensionality of the data set. However, it must be noted that the number of PCs needed to adequately explain original data varied a great deal and often appeared to be based on subjective expert opinion. PCA is able to contend with the large data sets that are customary in the field of gait analysis (Astephen *et al.*, 2002a; Astephen *et al.*, 2002b). Consequently, PCA is suitable as a pre-processing step subsequent to the application of other methods as seen in the study of Shiavi and Griffin (1981). Like FA, PCA has the ability to represent gait waveforms in a discrete form whilst retaining temporal information (Deluzio *et al.*, 1997; Deluzio *et al.*, 1999; Sadeghi *et al.*, 2000; Astephen *et al.*, 2002a; Astephen *et al.*, 2002b). Additionally, it is able to deal simultaneously with continuous and discrete variables (Astephen *et al.*, 2002a; Astephen *et al.*, 2002b). The transformed variables can have a physical meaning that is clinically relevant although this requires manual intervention and is dependent on the ability of the researcher to interpret the PCs. This interpretation stage is therefore a subjective and often time-consuming process and it is not always possible to assign a physical meaning to the PCs. If the PCs have a physical meaning, they can be used to detect changes in gait due to surgical or therapeutic intervention. This can be achieved in one of two ways. Firstly, if two PCs can adequately represent the original data, the

transformed data can be displayed in bivariate plots. This can enable changes to be represented visually (Yamamoto *et al.*, 1983). Secondly, the PCs can be utilised to form a gait index again capable of highlighting gait changes or differentiating distinct subject populations (Schutte *et al.*, 2000).

1.2.2.2 Linear Discriminant Analysis (LDA)

A fuller description of LDA is given in Chapters 5 and 7; consequently, a concise summary is given here. The objective of LDA is to obtain a weighted, linear combination of as set of variables such that the different groups of subjects are as distinct as possible. LDA yields a discriminant function of the form:

$$D = a + b_1 v_1 + b_2 v_2 + \ldots + b_p v_p \tag{1.4}$$

where D is the discriminant score, p the number of variables, a a scaling factor, b_i the discriminant function coefficient and v_i the discriminator variable.

Chao *et al.* (1980) employed LDA to compare the knee function of pre and post-operative TKR patients with NL subjects. Through examination of the weighting coefficients of the linear discriminant function, eight out of forty three parameterised gait variables were found to be most significant in providing discriminative power in separating the two subject groups. Using these eight variables, a performance index was developed to assess the overall functional status of a patient. From the functional indices, the group were able to identify differences between the NL and pathological subjects and changes due to surgical intervention. This study highlights the ability of LDA to identify the most important gait parameters when differentiating two subject groups and to reduce the number of gait variables used in subsequent analyses. Furthermore, the discriminant function can be evaluated for each subject to give a single value corresponding to knee function.

Subsequent to performing PCA on a combined data set of constant gait parameters and time-varying gait waveforms, Astephen *et al.* (2002a; 2002b) employed a stepwise linear discrimination procedure to identify the PCs most able to discriminate between OA and NL gait patterns. These features were interpreted through

examination of the contribution of each of the input variables to each feature. This study revealed that the loading response phase of the gait cycle is an important factor in isolating OA.

LDA is able to indicate the gait parameters most able to discriminate different subject groups (Chao *et al.*, 1980; Astephen *et al.*, 2002a; Astephen *et al.*, 2002b). This inherently enables a reduction in the number of gait variables used in subsequent analyses. Furthermore, evaluating the discriminant function for each subject gives a single value corresponding to knee function (Chao *et al.*, 1980).

1.2.2.3 Cluster Analysis (CA)

"CA is the art of finding groups in data." (Kaufman and Rousseeuw, 1990) The aim of CA is to automatically group objects in a way that those in the same group are as similar as possible and those in different groups are as distinct as possible. Unlike LDA, CA establishes the groups itself and does not define them *a priori*. A number of different clustering algorithms exist and their use will depend on the type of data being used and the type of clusters required.

Subsequent to performing PCA on their data set to reduce its dimensionality, Shiavi and Griffin (1981) implemented a CA in order to detect groupings in their data. Using three PCs, the patterns were split into five clusters, which were found to be statistically different. With three PCs, CA identified prevalent muscle patterns although little interpretation was given of these patterns. This early work revealed the potential of CA for pattern detection.

1.2.3 The Artificial Neural Network (ANN)

The interest in, and use of, ANNs has grown rapidly over recent years primarily due to their inherent non-linear modelling ability. This is reflected in the vast amount of research that has been executed in the area of gait analysis since its first application in this field in 1993. Indeed, ANN dominate the types of analysis techniques used in gait analysis over the last decade. The types of ANN employed within these studies are shown in Figure 1.6

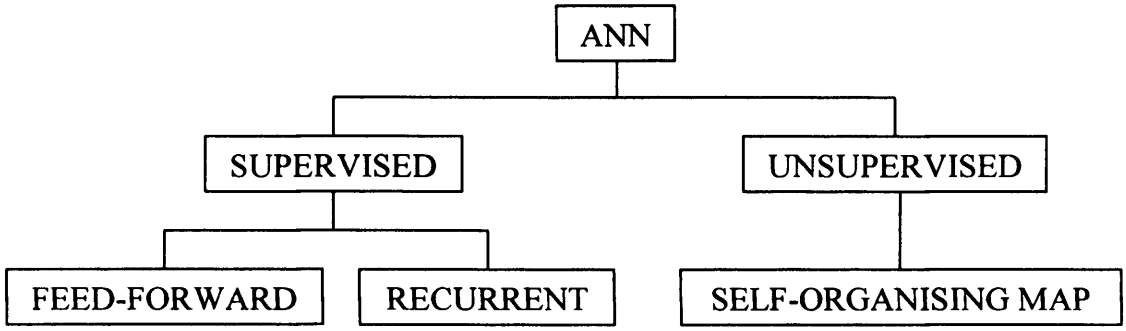


Figure 1.6 ANN employed within gait analysis studies.

An in-depth description of the ANN is given in Chapter 7 (section 7.1); consequently, only a brief summary is given here.

The ANN consists of non-linear processing elements, neurons, which are grouped together in layers. Generally, the ANN comprises an input layer in which neurons are connected to the input data; one or more hidden layers, which are the processing layers; and an output layer in which neurons transmit the final output to the user.

Each neuron receives a set of inputs via weighted connections, weights. The neuron is activated if the sum of the weighted inputs exceeds the threshold value of the neuron. The signal is subsequently passed through an activation function, which produces the neuron output.

The way in which neurons are connected depends on the type of ANN being used. In a feed-forward ANN, neurons on one layer pass their output to the neurons on the subsequent layers but not to preceding layers. Thus, information passes through the network in a forward-direction only. In a recurrent ANN, each individual neuron passes its outputs to all the other neurons within that layer. This is performed several times until some condition is met. Only then does each neuron send its output to the subsequent layer.

The relationship between the input neurons and the output neurons is not intrinsic to the ANN but must be learnt through a process called training. Once the ANN has been trained, it can accept previously unseen data and attempt to predict an associated

output, simply by evaluating the internal functions. During supervised training, examples of inputs and their corresponding outputs are presented to the ANN. By iteratively self-adjusting the connection weights, the ANN learns to infer the relationship between the inputs and outputs. Training is achieved using a supervised learning algorithm e.g. the back propagation algorithm (see Bishop, 1995). In contrast, during unsupervised training only examples of inputs are presented to the network. The ANN attempts to learn the structure of the data and as a result can start to recognise clusters of data and relate similar classes to each other.

In this section, a review of the development of ANN for gait data analysis has been limited to studies concerned with the classification of human movement. For a review of studies concerned with modelling and prediction the reader is directed to Chau (2001b).

Holzreiter and Köhle (1993) trained a three-layer feed-forward, back propagation ANN to distinguish between pathological and NL gait. The training set comprised pre-processed vertical GRFs taken from 94 NLs and 131 patients (calcaneus fracture, prosthesis users). A total of 8173 input patterns were available, given that multiple trials were recorded for each subject. By means of randomly presenting these patterns of the training set, the ANN was trained using 200,000 iterations and then a test set was applied to evaluate the ability of the ANN to correctly classify previously unseen samples. The rate of success using the test set was between 75 and 95%. In addition to classifying input patterns, the ANN output included an estimate of how well a pattern fitted into its assigned class. This enabled a more subtle classification of input patterns.

Barton and Lees (1995) developed an ANN to classify patients with foot abnormalities. Dynamic pressure patterns, comprising 1316 pressure values, were recorded for 18 subjects. A four-layer feed-forward, back-propagation ANN was constructed. Despite use of a relatively small training set, Barton and Lees (1995) reported accuracies in the range of 77-92% in classifying new data. However, testing was performed with a very small number of patterns, and the relatively high classification accuracies must be viewed cautiously. Furthermore, the ANN achieved higher accuracy at the expense of a simple structure.

Gioftsos and Grieve (1995) constructed three-layer recurrent ANN to discriminate patterns of human gait. Their performance was compared with that of LDA. Temporal parameters of twenty NL subjects were recorded whilst walking at different speeds and under different conditions (NL walking, walking with a mass attached to the ankle and walking with the right knee fixed in a brace). ANNs were used for three different applications:

1. Recognition of walking speed in a given walking condition
2. Recognition of walking condition at a given speed
3. Recognition of walking condition.

The training set lacked sufficient samples, which resulted in over-training; only ten patterns were used for the training of each ANN. Up to 73% accuracy was achieved in correctly recognizing unknown patterns, compared to 68% when using LDA. On this basis, Gioftsos and Grieve could not conclude that using ANN was significantly better than LDA. This study highlights the necessity of using sufficiently large training set consisting of a wide range of possible categories.

Köhle and Merkl (1996) employed self-organising maps (SOM) to identify various pathological gait patterns based on GRFs. The pathologies were associated with the right pelvis; right and left pelvis (pelvis both sides); artificial left ankle; artificial right ankle; left knee and right foot; right knee; left foot and additional injuries; left foot and left ankle; and the left ankle. The SOM comprises a grid of artificial neurons each having a physical location on an output map. During unsupervised training, the weights of these neurons are adapted to match the input vectors in the training set. This results in similar input vectors being clustered together on the output map. New subjects are then assigned to a location on the output map that is most representative of their input pattern. The clustering of similar input vectors is evident in the work of Köhle and Merkl since patients with similar pathologies are assigned to neighbouring areas of the output map as shown in Figure 1.7. The classification of subjects was performed automatically i.e. it was not necessary to assign a label to each gait pattern before training the SOM. Thus, the gait patterns of patients were classified solely using GRFs.

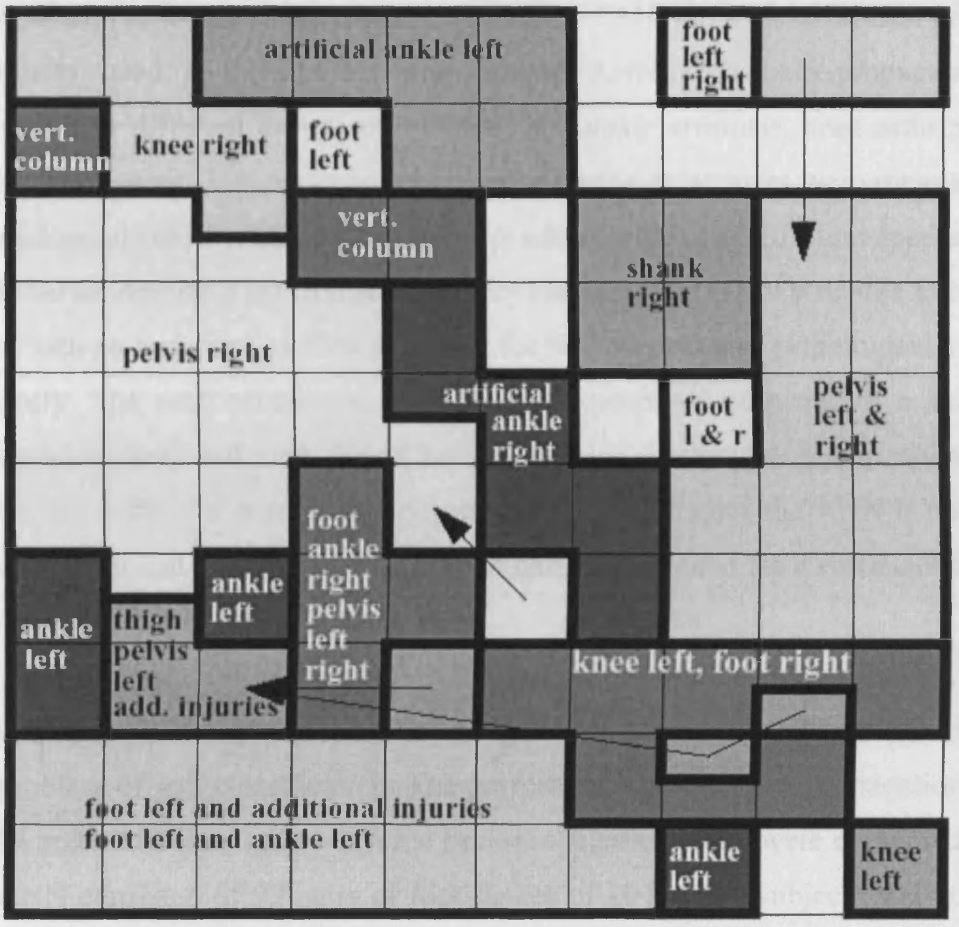


Figure 1.7 Map of gait patterns from 100 patients. The different shading of output units identifies the varying gait pathologies. The arrows connect areas of similar pathologies that are located in separate parts of the map. (Taken from Köhle and Merkl, 1996, Figure 4).

Barton and Lees (1997) utilised ANN for distinguishing gait patterns based on hip-knee joint angle diagrams. The motion of eight subjects was recorded whilst walking under three different conditions, namely NL walking, a simulation of leg length difference and a simulation of leg weight asymmetry. Pre-processing of data involved time normalisation, fast Fourier transformation and linear transformation. A four-layer, feed-forward, back-propagation ANN was constructed, although no justification was given for the choice of two hidden layers. Outputs of the ANN were the three different gait conditions. Training was performed with the use of 18 patterns (eight people), and testing with the use of six patterns (two people). Subsequent to training, the ANN was able to correctly identify new patterns with an average accuracy of 75%. Once more, only a small testing and training set was utilised.

Lafuente *et al.* (1998) investigated the application of ANN to a classification problem. The authors used a three-layer, feed-forward ANN with back-propagation, to distinguish four different groups of subjects: NL, ankle arthrosis, knee arthrosis and hip arthrosis. Kinetic, kinematic and temporal-distance parameters were measured for 148 pathological subjects and 88 NL subjects whilst walking at different speeds. LDA was utilised to identify a set of discriminatory features. The ANN was able to classify subjects with an accuracy of 87% and 73% for NL subjects and pathological subjects respectively. The performance of the ANN was compared with that of a statistical Bayesian classifier, and was found to be superior at the 5% significance level. However, the authors comment that despite satisfactory results, the ANN is not ready for practical, clinical use. Again, this study highlights the need for a sufficient training sample size.

Su and Wu (2000) introduced the genetic algorithm (GA) ANN (GAANN) approach to the problem of gait classification. The performance of the back-propagation based GAANN and a traditional three-layered back-propagation ANN were compared. Input to the ANN consisted of 99 pairs of foot strikes of 10 healthy subjects and 10 ankle arthrodesis patients. Subsequent to training, the GAANN was able to correctly classify subjects to an accuracy of 98.7% whilst the ANN had an accuracy of 89.7%. The study suggests that ANN using GA are able to classify subjects with a higher accuracy rate. Furthermore, Su and Wu (2000) used the value of the neuron output as an index of the difference from NL and thus represented pathological gait patterns quantitatively. This study demonstrates that GAANN are a powerful tool for discriminating between NL and pathological subjects.

Wu and Su (2000) developed a three-layered back-propagation ANN to distinguish between gait patterns of NL subjects and ankle arthrodesis patients based on a set of force parameters. A stepwise discriminant procedure was used to reduce the number of parameters that were used as inputs to the ANN. The ANN, with a simple structure (5 hidden neurons) was able to classify subjects with accuracies of up to 95.8%. The performance of the ANN was compared with that of LDA, which was able to classify subjects with an accuracy of 91.5%.

ANN have been applied extensively to studies of NL and pathological gait. From the work of Barton and Lees (1995), it is evident that ANN have the ability to deal with the huge amount of data intrinsic to most gait studies. In many studies, ANN have been developed as classification tools, although their success in such a task has varied greatly. It is evident from the reviewed studies that the success of ANN is highly susceptible to the size of the training set. The inadequacy of training data is a common problem as shown in the studies of Gioftsos and Grieve (1995) and Lafuente *et al.* (1998). This has led to over-training and poor generalisation ability. The studies with high reported classification accuracies required vast training sets as seen in the study of Holzreiter and Köhle (1993) or used very small testing sets (Barton and Lees, 1995, 1997). Furthermore, a successful performance of ANN is dependent on the appropriate selection and pre-processing of input variables: a choice that is not straightforward and is dependent on experience and expert opinion. Studies in the literature used a variety of pre-processing techniques including scaling, normalisation and Fast-Fourier transformation. The inclusion of GA can greatly enhance the accuracy of ANN as seen in the work of Su and Wu (2000).

1.2.4 Fuzzy methods

A small number of studies have used Fuzzy methods to analyse gait data. The main thrust of their work has been to detect natural groupings among subjects and then to use these groupings to classify new patients and identify changes due to surgical intervention.

In classic set theory, objects either completely belong to a set or they are completely excluded from it; it is not possible for an object to partially belong to a set. For example, considering a set of people with OA knee function, in classic set theory the boundary between OA and NL has to be set. If the boundary were set at osteoarthritis present in 75% of the knee, a person with 76% presence would belong to the "OA knee function" set whereas a person with 74% would not, as highlighted in Figure 1.8. This system does not describe realistically the transition from NL to OA knee function.

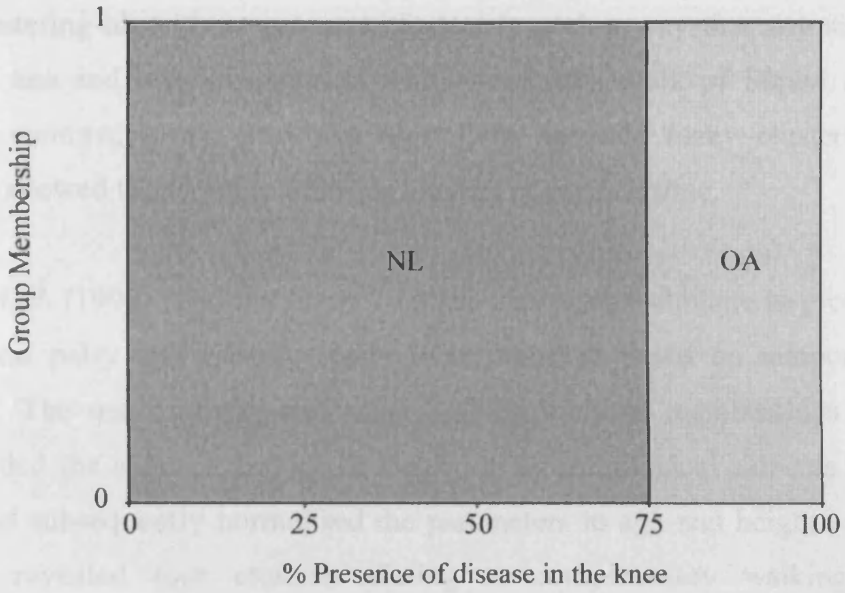


Figure 1.8 Classic set theory.

Fuzzy set theory, first introduced by Zadeh (1965) is an extension of classic set theory, in which an object is allowed to belong partially to two mutually exclusive sets. For example, as shown in Figure 1.9 a person with 75% presence of OA can have full membership of the OA set, whereas the one with 74% can have 95% membership of the OA set and 5% membership of the NL set, say. In this way, the question "to what extent does a person have OA knee function?" can be answered and categories such as "severely OA", "mildly OA" can be defined.

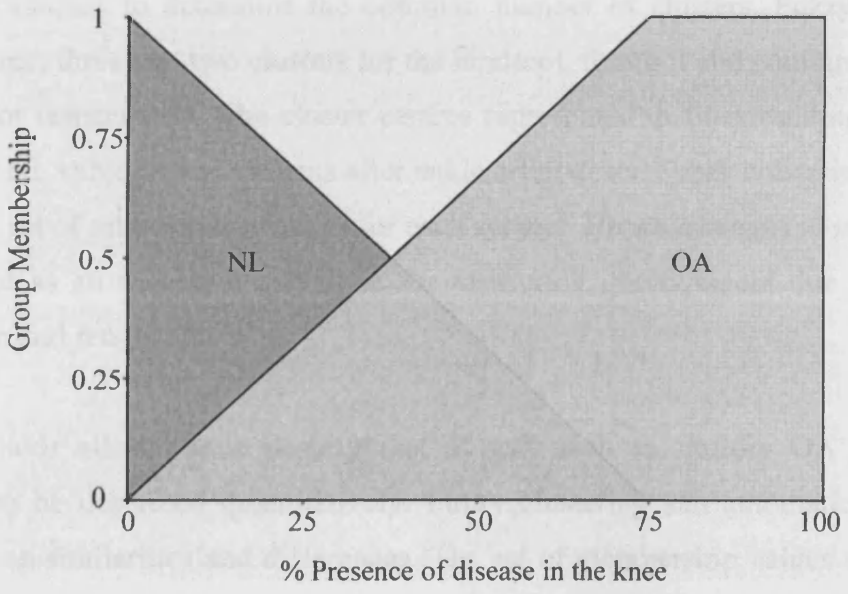


Figure 1.9 Fuzzy set theory.

Classic clustering algorithms generate clusters in such a way that any single object belongs to one and only one cluster, as shown in the work of Shiavi and Griffin (1981). In contrast, fuzzy clustering algorithms generate fuzzy clusters in which objects are allowed to belong to multiple clusters at any one time.

O'Malley *et al.* (1997) used the fuzzy k-means clustering technique to group children with cerebral palsy and neurologically intact children based on temporal-distance parameters. The use of fuzzy-clustering enabled multiple memberships of groups. They recorded the cadence and stride length of 88 pathological subjects and 68 NL subjects and subsequently normalised the parameters to age and height respectively. Clustering revealed four clusters relating to compensatory walking strategies associated with cerebral palsy, and one cluster representing neurologically intact gait. Subsequently, pre and post-operative test data, not part of the original data set was introduced. Membership values for the five clusters were calculated for each individual belonging to the test set. Improvement due to surgical intervention was objectively identifiable in terms of changes in the cluster membership values.

Su *et al.* (2001) investigated the use of the k-means fuzzy clustering algorithm as a means of classifying pathological and NL gait patterns. Sagittal plane Euler angles for the hindfoot, forefoot and combined hindfoot and forefoot joints were recorded for 10 NL subjects and 10 patients with ankle arthrodesis. Su *et al.* (2001) used cluster validity techniques to determine the optimum number of clusters. Fuzzy clustering revealed three, three and two clusters for the hindfoot, forefoot and combined forefoot and hindfoot respectively. The cluster centres represented distinct walking strategies adopted by NL subjects and patients after ankle arthrodesis. Fuzzy clustering yields as an output a set of membership values for each subject. Hence, changes in membership can be used as an objective technique for measuring improvement due to surgical intervention and recovery.

Fuzzy methods allow vague descriptions of gait such as 'mildly OA', 'severely impaired' to be described quantitatively. Fuzzy clustering can automatically group data based on similarities and differences. The set of membership values assigned to each subject can be used to objectively measure changes in gait function (Su *et al.*, 2001; O'Malley *et al.*, 1997). Fuzzy methods cannot deal directly with time-varying

data; it can only deal with data that has been parameterised into an appropriate set of features. Finally, certain parameters have to be chosen *a priori*, which means that the method relies heavily on expertise.

1.2.5 Discussion

Past research has shown that FA, PCA, LDA, CA, fuzzy methods and ANN have much to offer to the analysis of gait data. The main issues regarding their application are now discussed.

The success of many of the methods to analyse gait data is highly dependent on the appropriate selection, pre-processing and reduction of input variables. The selection process, in the main, has depended highly on expert opinion. Although feature selection methods exist (see Dash and Lui, 1997; Siedlecki and Sklansky, 1988), only one of the above studies took advantage of them. The work of Su and Wu (2000) highlighted the benefits of employing such feature selection techniques: their use of a GA improved the accuracy of the ANN by 9%. The majority of studies necessitated pre-processing of input variables. Both FA and PCA have proved to be effective pre-processing methods. Each technique is capable of providing reduced representations of continuous data whilst retaining temporal information. However, of the two techniques, PCA can produce representations that have a physical and clinically relevant meaning. Additionally, the success of powerful techniques such as ANN is further dependent on the size of the training set used. The most successful applications used training sets comprising 100 to 8000 input patterns, which is often not practically attainable in the clinical setting.

When using PCA, LDA and ANN it is possible to identify the individual contribution each variable makes to the overall output. In the cases of PCA and LDA this contribution is evident in the weighting factors. For ANN, the weights and biases of the ANN capture the structure of the relationship between the input variables and the outcome. However, these are difficult to interpret and none of the above studies has attempted this.

The PC loadings, linear discriminant functions, ANN neuron output and fuzzy membership values, can be utilised to give a single value or set of values corresponding to knee function. As a result, a gait index can be established and used to compare NL and pathological gait and to detect and quantify changes due to surgical intervention. The visualisation of results can aid in the clinical interpretation of results and compliment the use of a gait index in the comparison of knee function and identification of changes due to surgical intervention. This is particularly evident from the binary plot used by Yamamoto *et al.* (1983) and the SOM utilised by Köhle and Merkle (1996). However, the use of binary plots limits comparison to two variables, and the SOM only shows the final classification of a subject and not the contribution the input variables make to this classification. The visualisation of results continues to be a challenge.

In different ways and to varying degrees, each method is capable of distinguishing different gait patterns. However, it is evident that no one technique can independently meet all the necessary requirements discussed in section 1.1.2. It has become apparent that a combinatorial approach will be essential to solve this problem (Chau, 2001b).

1.3 INTRODUCTION OF NEW METHOD

This work introduces a new classification method that incorporates the Dempster-Shafer theory of evidence (DST). Founded on the work of Dempster (1968) and Shafer (1976), the DST is a method that enables decision-making in the presence of uncertain, inadequate and conflicting evidence, a common problem in the CMA laboratory.

Safranek *et al.* (1990) and Gerig *et al.* (2000) utilised aspects of the DST in studies in the areas of vision and medical image analysis respectively and their works highlight the potential of the DST as a classification tool. In response to these findings, a preliminary study was conducted to investigate the efficacy of the DST in classifying gait data (Beynon *et al.*, 2002). An introduction to a novel method (the DS method or DS classifier) that incorporated the DST, an optimisation technique and simplex plots was introduced to classify NL and OA knee function. Exploiting the DST, the approach allowed for a degree of ignorance in the subject's classification, i.e., a level

of uncertainty as to whether or not a gait variable indicates normality. The control variables governing the transformation from input variables (ranges of motion in the three planes, cadence and peak vertical force) to final classification were indicative of the variables most influencing the final classification. These control variables were optimised using a simulated annealing algorithm, and their magnitude may provide useful information as to which are the most significant variables involved in the analysis of knee function. The inclusion of simplex plots allowed the classification of the subject and each associated characteristic to be represented visually. The DS method was used initially as a classification tool for two subjects and both were classified correctly.

Building on the work of Beynon *et al.* (2002) the remainder of this thesis investigates the use of the DS method in the classification of subjects with OA and NL knee function. The following chapters are based upon a number of key objectives:

1. Develop a method that is able to distinguish OA and NL knee function based on the measurements taken during knee clinical trial

2. Describe kinematic and kinetic waveforms in a discrete form without discarding temporal information

3. Ensure that the method is valid, accurate, comparable with other classification methods and clinically relevant.

4. Incorporate visualisation and highlight the contribution that each variable makes to the overall classification to enhance clinical interpretation

5. Decide which are the most important variables to input to the method and investigate whether an automated or expert approach should be used

6. Quantify changes following surgical intervention

1.4 THESIS SUMMARY

Chapter 2 introduces the motion analysis methods that were used to measure the knee function of OA, NL and TKR subjects, in terms of kinematic and kinetic waveforms, temporal-distance parameters, anthropometrical parameters and other patient related information. The method used to deal with the temporal waveforms is subsequently described. **Chapter 3** provides a detailed description of the DS method. The chapter places the DS method in its context and then explains, in detail, its application to the classification of OA and NL knee function. **Chapter 4** reports the results obtained from a series of tests that were conducted to investigate the validity of the DS classifier. **Chapter 5** presents a study which examines the use of feature selection. The study builds on a paper that was presented at the 1st International Congress in Computational Bioengineering (Jones *et al.*, 2003b). **Chapter 6** explores the potential of the DS classifier as a tool for assessing the outcome of TKR surgery. **Chapter 7** compares the performance of the DS classifier with two other, well-established classifiers. The chapter builds on work that was presented in two conference papers. The first paper was given as a podium presentation at the 3rd Lower Limb Conference and was awarded the "Best Abstract Award" (Jones *et al.*, 2003a). The second paper was presented at the 6th International Conference in Biomechanics and Biomedical Engineering (Jones *et al.*, 2004). **Chapter 8** provides a set of conclusions which can be drawn from this work and provides directions for future work.

CHAPTER 2

EXPERIMENTAL PROCEDURE

This chapter describes in detail the methods used to assess the knee function of normal (NL), osteoarthritic (OA) and total knee replacement (TKR) subjects. The raw data used in this work was collated as part of an on-going clinical trial conducted in the motion analysis laboratory in the School of Engineering, Cardiff University. The protocol for data collection is described in section 2.1. Subsequent to motion capture, the raw data was processed to produce a dataset of information relating to a subject's knee function as described in sections 2.2 to 2.4. The kinematic and kinetic waveforms produced in sections 2.2 and 2.3 respectively were further processed using Principal Component Analysis as described in section 2.5. Finally, a database was created for ease of data storage as shown in section 2.6.

2.1 DATA COLLECTION

The raw data was collected as part of a clinical trial established to assess the knee function of NL, OA and TKR subjects. For the work contained in this thesis, the raw data of a sample of 22 OA subjects (the OA sample), 20 NL subjects (the NL sample) and 9 TKR subjects (the TKR sample) was considered. The clinical trial was conducted in the motion analysis laboratory in the School of Engineering, Cardiff University. The laboratory is equipped with an opto-electronic measurement system (MacReflex system, Qualisys Inc.) comprising five digital infrared cameras (Figure 2.1a), two force platforms (Bertec Corporation, Number N60202, Type 4060H) embedded in a raised walkway (Figure 2.1b), a computer and two video cameras.

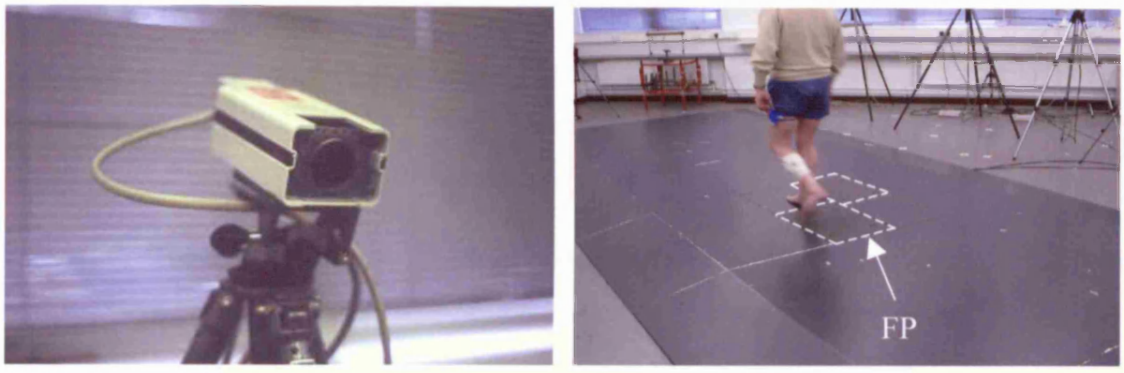
The clinical trial protocol for data collection was established by Holt *et al.* (2000). The set-up of the motion analysis laboratory during the clinical trial is shown in Figure 2.2. The infra-red cameras placed on one side of a walkway are positioned as to enable the detection of the movement of markers attached to the lower limbs on one side of the body. One video camera is located at the end of the walkway giving a front-on view of a subject's gait, and the other at the side of the walkway giving a

side-on view. Prior to a clinical trial subject visit, a calibration procedure is performed to establish the global coordinate system (GCS) within the laboratory. A calibration frame (Figure 2.3a) is placed on the floor of the laboratory and a marked wand (Figure 2.3b) is moved over the frame to calibrate a bounding volume large enough to capture the movement of a subject's legs during one complete gait cycle.

A clinical trial subject visit then progresses in a number of stages:

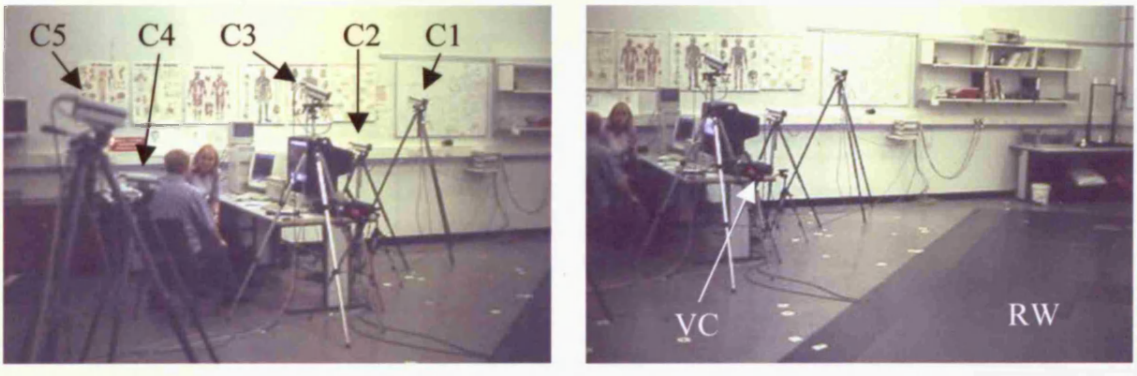
1. Collection of subject details
2. Anthropometrical measurements
3. Placement of marker clusters on shank and thigh
4. Identification of bony landmarks by palpation
5. Neutral position measurement
6. Walking trials
7. Knee outcome survey

These stages, which are completed in turn for each leg, will now be described in more detail.



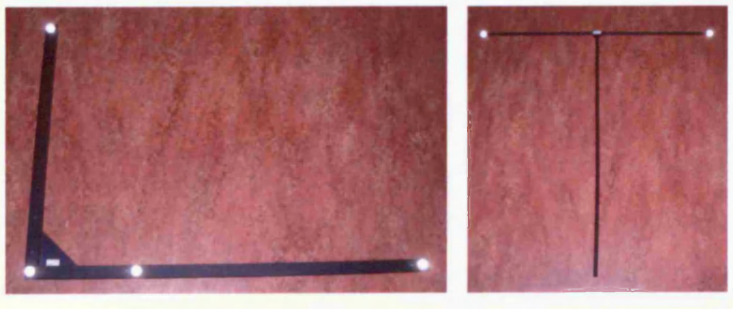
(a) (b)

Figure 2.1 (a) The infra-red camera and (b) the force platforms (FP) (highlighted by dotted lines) embedded in the raised walkway.



(a) (b)

Figure 2.2 The set up of the motion analysis laboratory during the clinical trial. (a) The opto-electronic measurement system comprising five infra-red cameras (C1 to C5) are (b) positioned along one side of the raised walkway (RW) along with a video camera (VC).



(a) (b)

Figure 2.3 a) Calibration frame and b) calibration wand.

2.1.1 Collection of Subject Details

At the start of the clinical trial visit, a set of subject specific details is collected as listed in Table 2.1.

Table 2.1 Subject details recorded during the clinical trial visit.

Name
Address
Telephone Number
Age
Date of Birth
Sex
Leg
Date of first assessment by doctor (if applicable)
Operation date (if applicable)
Surgeon (if applicable)
Knee type (OA/NL/TKR)
Date of trial
Visit type (pre-op/3 months post-op/6 months post-op/12 months post-op/normal cohort)
Subject history – injuries/disabilities
Subject comments

2.1.2 Anthropometrical Measurements

Subsequently, a number of anthropometrical measurements are taken: height, h (m), weight, w (kg), anterior-posterior knee width (cm), medial-lateral knee width (cm), and thigh girth (cm). A subject's height and thigh girth are measured using a tape measure, their weight with weighing scales and their knee widths with measuring

callipers. The thigh girth measurements are repeated three times and an average value is recorded.

2.1.3 Placement of Marker Clusters on Shank and Thigh

Passive four-marker clusters were previously made using copolymer polypropylene and covered with retro-reflective markers as shown in Figure 2.4. These marker clusters are positioned laterally on the shank and thigh and held in position using self-adhesive Coban© tape or tubigrip as shown in Figure 2.5. The marker clusters remain in position throughout the remainder of the trial. The use of marker clusters rather than single markers and their placement at sites known to experience minimal overlying skin movement, minimises skin movement artefacts associated with standard motion analysis techniques (Cappello *et al.*, 1997).

2.1.4 Identification of Bony Landmarks by Palpation

During an initial calibration with the subject standing in a neutral position, the point of a marked pointer is used to identify the 3-D coordinates of three bony landmarks per segment (shank and thigh) as shown in Figure 2.6. Using manual palpation, each bony landmark is identified in turn for the shank (medial condyle, lateral condyle and medial malleolus) and thigh (medial condyle, lateral condyle, upper border of the trochanter) with the marker clusters attached to the upper and lower limb segments. The location of the bony landmarks is shown in Figure 2.7. During these initial calibrations, the opto-electronic measurement system locates the positions of the marker clusters and pointer in terms of the GCS of the laboratory. Sixty frames of coordinate data are measured during a one second data collection period.

2.1.5 Neutral Position Measurement

With the subject standing in a neutral position as shown in Figure 2.5, the opto-electronic measurement system measures the positions of the marker clusters attached to the upper and lower limb segments in terms of the GCS of the laboratory, for a one second data collection period.

Figure 2.4 A rigid marker cluster covered with four retro-reflective markers.

Figure 2.5 The marker clusters are positioned laterally on the shank and thigh and are held in position using self-adhesive tape or tubigrip.

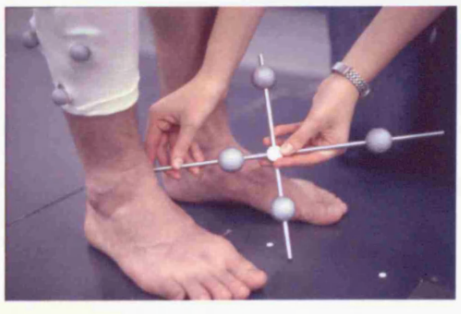


Figure 2.6 Identification of bony landmarks using the point of a marked pointer, covered in retro-reflective markers.

2.1.6 Walking Trial

The opto-electronic measurement system records the movement of the retro-reflective marker clusters attached to the shank and thigh and the force platform records the ground reaction forces (GRF) as the subject walks along a raised walkway, as shown in Figure 2.8. During each trial, the motion data and force plate readings are acquired for a single complete gait cycle with sampling frequencies of 60Hz and between 960 and 1080 Hz respectively. The length of the walkway allows the subject to take four or five strides before any measurements are taken. The subject is lined up before walking to maximise their potential of hitting the force-plate (although they are unaware of the force plate as a target). The subject is allowed to rest between each recording if they become tired. Six successful walking trials are recorded for each subject as they walk at their usual walking pace. A trial is deemed successful if the subject cleanly contacts the force plate. The two video cameras provide a synchronized visual record of the subject walking.

2.1.7 Knee Outcome Survey

A subject independently completes The Activities of Daily Living Scale of the Knee Outcome Survey (KOS) (Irrgang *et al.*, 1998). The KOS is a subjective measure of the symptoms and functional limitations, experienced by a subject during daily activities, resulting from their knee pathology. KOS consists of seventeen questions with seven relating to symptoms and ten to functional disability. Each question is followed by a number of associated statements and the subject must tick the statement which best describes their recent experience. A scoring system is used to assign values to each question. The final score is subsequently displayed as a percentage, with a high final score being associated with a high level of function and vice-versa. The KOS may be used to assess a subject's function at a particular instance in time or to assess changes in their function over a longer period (Irrgang *et al.*, 1998). A copy of the questionnaire is given in Appendix A.

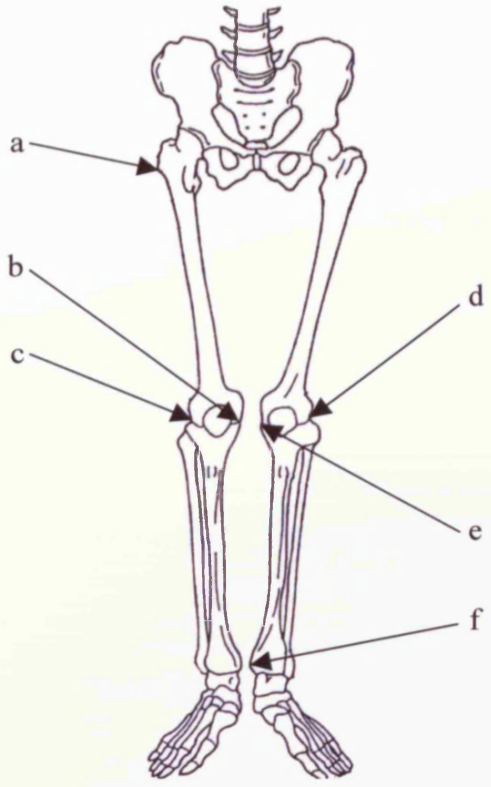


Figure 2.7 Location of bony landmarks for the femur: a) upper border of the trochanter, b) medial condyle, c) lateral condyle; and the tibia: d) lateral condyle, e) medial condyle, f) medial malleolus. (Adapted from Whittle, 1996, pp.6)

Figure 2.8 A walking trial showing the subject walking along the raised walkway and cleanly contacting the force platform (highlighted by the dotted white line) in the stance phase.

2.2 DATA PROCESSING - KINEMATICS

QTrac software (Qualysis Inc.) was used to track the movement of the marker cluster and pointer markers into 3-D GCS coordinates for the anatomical calibrations, neutral position and walking trials of the OA sample, the NL sample and the TKR sample. Data corresponding to one gait cycle was tracked for the walking trials. Previously developed Matlab (version 11.1, The MathWorks Inc.) software (Holt *et al.*, 2000) was applied to the coordinate data of all three samples to produce the knee joint rotations and translations for one gait cycle. The methods used in this software are summarised in the following steps:

For the anatomical calibration, for each bony landmark:

1. Establish a pointer local coordinate system (PLCS) using the GCS coordinates of the pointer markers as shown in section 2.2.1.1.

2. Calculate the transformation matrix $\mathbf{T}_{pg}$ relating the orientation of the PLCS in the GCS as shown in section 2.2.2.

3. Establish a marker cluster local coordinate system (MLCS) using the GCS coordinates of the marker cluster markers as shown in section 2.2.1.2.

4. Calculate the transformation matrix $\mathbf{T}_{mg}$ relating the orientation of the MLCS in the GCS as shown in section 2.2.2.

5. Calculate the MLCS coordinates of the pointer's point (P_m) using (2.1). This establishes the location of the bony landmark in the MLCS:

$$\mathbf{P}_m = [\mathbf{T}_{gm}][\mathbf{T}_{pg}][\mathbf{P}_p] \qquad (2.1)$$

where P_p are the PLCS coordinates of the pointer's point and $\mathbf{T}_{gm}$ is the inverse of $\mathbf{T}_{mg}$.

For the anatomical calibration, for each segment:

6. Establish an anatomical local coordinate system (ALCS) as a fixed body axis on the bone, using the MLCS coordinates of the three bony landmarks (P_m) on that segment as shown in section 2.2.1.3 for the tibia and 2.2.1.4 for the femur.

7. Calculate the transformation matrix $\mathbf{T}_{ma}$ relating the orientation of the MLCS in the ALCS as shown in section 2.2.2.

For the neutral position or walking trial:

8. Recalculate $\mathbf{T}_{mg}$ for the shank and thigh (see point 4 above)

9. Calculate the transformation matrix $\mathbf{T}_{tf}$ relating the orientation of the tibial ALCS to the femoral ALCS using (2.2)

$$\mathbf{T}_{tf} = [\mathbf{T}_{ma,T}][\mathbf{T}_{gm,T}][\mathbf{T}_{mg,S}][\mathbf{T}_{am,S}] \qquad (2.2)$$

where the subscripts T and S refer to the thigh and shank segments respectively, and $\mathbf{T}_{gm}$ and $\mathbf{T}_{am}$ are the inverses of $\mathbf{T}_{mg}$ and $\mathbf{T}_{ma}$ respectively.

10. Apply the Joint Coordinate System (JCS) approach (Grood and Suntay, 1983) to $\mathbf{T}_{tf}$ to obtain the flexion-extension, abduction-adduction and internal-external rotation angles of the tibio-femoral joint as shown in section 2.2.3.

11. Repeat steps 8 to 10 for each frame of the gait cycle

12. The three resulting rotation waveforms are re-sampled over 100 points and an average of the six walking trials for each waveform is computed.

2.2.1 Establishing a local coordinate system on a rigid body

To establish a local coordinate system (LCS) on a rigid body, the coordinates of at least three non-collinear positions on that body must be known (Zatsiorsky, 1998). The coordinates of these three (or more) positions are utilised to establish three mutually orthogonal vectors, which form the LCS (Zatsiorsky, 1998). The procedure

used to establish the PLCS, MLCS, tibial ALCS and femoral ALCS are now discussed.

2.2.1.1 Establishing the PLCS

To define three mutually orthogonal vectors and establish the PLCS the following protocol is followed in the software developed by Holt *et al.* (2000):

1. The origin of the PLCS is located in marker P_2 as shown in Figure 2.9.
2. A vector $\mathbf{U}_1$ is defined from P_2 to P_1
3. A vector $\mathbf{U}_2$ is defined from P_3 to P_4
4. The cross product of $\mathbf{U}_1$ and $\mathbf{U}_2$ gives the vector $\mathbf{U}_3$
5. The cross product of $\mathbf{U}_1$ and $\mathbf{U}_3$ gives the vector $\mathbf{U}_4$
6. Vectors $\mathbf{U}_4$, $\mathbf{U}_1$, $\mathbf{U}_3$ divided by their own lengths give the unit vectors $\mathbf{u}_4$, $\mathbf{u}_1$ and $\mathbf{u}_3$ which produce the x, y, z axes respectively, attached to the origin at P_2.

2.2.1.2 Establishing the MLCS

To define three mutually orthogonal vectors and establish the MLCS the following protocol is followed in the software developed by Holt *et al.* (2000):

1. The origin of the MLCS is located in marker M_2 as shown in Figure 2.10.
2. A vector $\mathbf{V}_1$ is defined from M_2 to M_1
3. A vector $\mathbf{V}_2$ is defined from M_3 to M_4
4. The cross product of $\mathbf{V}_1$ and $\mathbf{V}_2$ gives the vector $\mathbf{V}_3$
5. The cross product of $\mathbf{V}_1$ and $\mathbf{V}_3$ gives the vector $\mathbf{V}_4$
6. Vectors $\mathbf{V}_4$, $\mathbf{V}_1$, $\mathbf{V}_3$ divided by their own lengths give the unit vectors $\mathbf{v}_4$, $\mathbf{v}_1$ and $\mathbf{v}_3$ which produce the x, y, z axes respectively, attached to the origin at M_2.

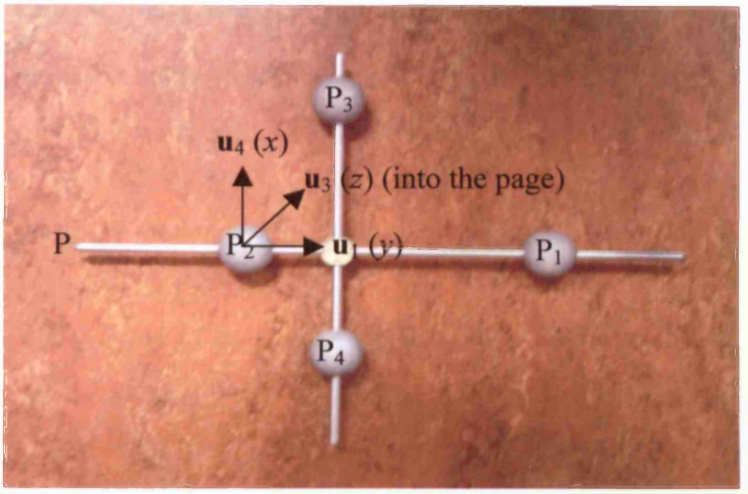


Figure 2.9 The PLCS is established using the four pointer markers P_1 (right), P_2 (left), P_3 (top) and P_4 (bottom). P denotes the point of the pointer. The unit vectors $\mathbf{u}_4$, $\mathbf{u}_1$ and $\mathbf{u}_3$ produce the x, y, z axes respectively attached to the origin at P_2.

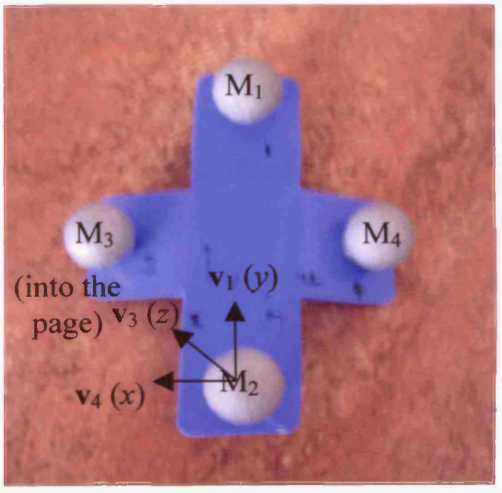


Figure 2.10 The MLCS is established using the four markers M_1 (top), M_2 (bottom), M_3 (left) and M_4 (right). The unit vectors $\mathbf{v}_4$, $\mathbf{v}_1$ and $\mathbf{v}_3$ produce the x, y, z axes respectively attached to the origin at M_2.

2.2.1.3 Establishing the tibial ALCS

To define three mutually orthogonal vectors to establish the tibial ALCS the following protocol is followed in the software developed by Holt *et al.* (2000):

1. The origin of the tibial ALCS is located at a point that lies half-way between the medial and lateral condyles
2. A vector $\mathbf{W}_1$ is defined from the medial condyle to the lateral condyle
3. A vector $\mathbf{W}_2$ is defined from the medial malleolus to the medial condyle
4. The cross product of $\mathbf{W}_1$ and $\mathbf{W}_2$ gives the vector $\mathbf{W}_3$
5. The cross product of $\mathbf{W}_1$ and $\mathbf{W}_3$ gives the vector $\mathbf{W}_4$
6. Vectors $\mathbf{W}_1$, $\mathbf{W}_3$, $\mathbf{W}_4$ divided by their own lengths give the unit vectors $\mathbf{w}_1$, $\mathbf{w}_3$ and $\mathbf{w}_4$ which produce the *x*, *y*, *z* axes respectively, attached to the origin as shown in Figure 2.11a.

2.2.1.4 Establishing the femoral ALCS

To define three mutually orthogonal vectors to establish the femoral ALCS the following protocol is followed in the software developed by Holt *et al.* (2000):

1. The origin of the femoral ALCS is located at a point that lies half-way between the medial and lateral condyles
2. A vector $\mathbf{N}_1$ is defined from the medial condyle to the lateral condyle
3. A vector $\mathbf{N}_2$ is defined from the medial condyle to the upper border of the trochanter
4. The cross product of $\mathbf{N}_1$ and $\mathbf{N}_2$ gives the vector $\mathbf{N}_3$
5. The cross product of $\mathbf{N}_1$ and $\mathbf{N}_3$ gives the vector $\mathbf{N}_4$
6. Vectors $\mathbf{N}_1$, $\mathbf{N}_3$, $\mathbf{N}_4$ divided by their own lengths give the unit vectors $\mathbf{n}_1$, $\mathbf{n}_3$ and $\mathbf{n}_4$ which produce the *X*, *Y*, *Z* axes respectively, attached to the origin as shown in Figure 2.11b.

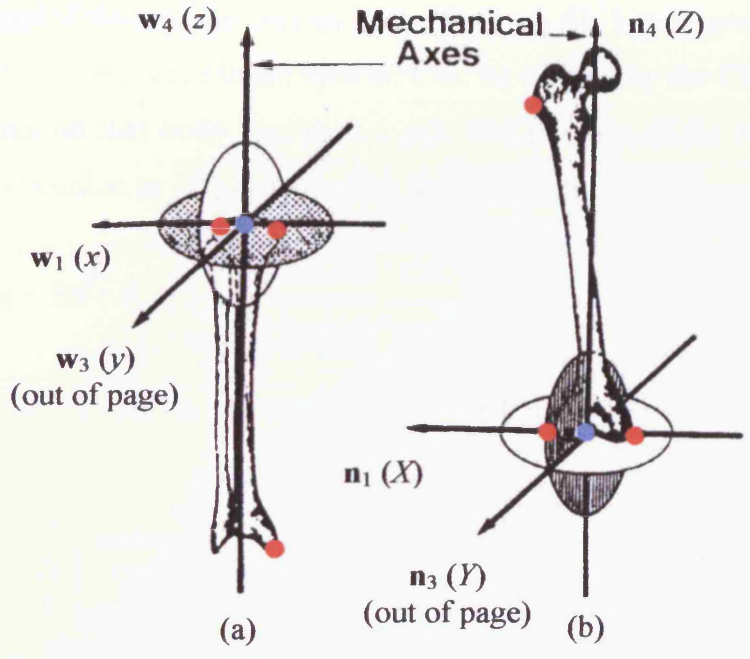


Figure 2.11 The coordinate systems defined in (a) the tibia and (b) the femur (adapted from Grood and Suntay, 1983, pp.138). Red circles identify bony landmarks. Blue circles denote origins of coordinate systems.

2.2.2 Defining the relative position of two coordinate systems

The software developed by Holt *et al.* (2000) uses the method of Söderkvist and Wedin (1993) to calculate the transformation matrices that define the relative position of two coordinate systems, namely $\mathbf{T}_{pg}$, $\mathbf{T}_{mg}$, and $\mathbf{T}_{am}$. This method will now be described.

Let the position of a rigid body in one coordinate system, CS_1, be defined by the CS_1 3D coordinates of three points on that body $\{f_1, f_2, ..., f_9\}$. Let the position of the same rigid body in a second coordinate system, CS_2, be defined by the CS_2 3D coordinates of three points on that body $\{g_1, g_2, ..., g_9\}$. The position of the rigid body in CS_1 relative to its position in CS_2 is defined by the mapping

$$\mathbf{g} = \mathbf{R}\mathbf{f} + \mathbf{d} \tag{2.3}$$

where $\mathbf{R}$ and $\mathbf{d}$ are a rotation matrix and translation vector respectively, $\mathbf{f} = \begin{bmatrix} f_1 & f_4 & f_7 \\ f_2 & f_5 & f_8 \\ f_3 & f_6 & f_9 \\ 1 & 1 & 1 \end{bmatrix}$ and $\mathbf{g} = \begin{bmatrix} g_1 & g_4 & g_7 \\ g_2 & g_5 & g_8 \\ g_3 & g_6 & g_9 \\ 1 & 1 & 1 \end{bmatrix}$.

or

$$\mathbf{g} = \mathbf{T}\mathbf{f} \tag{2.4a}$$

where

$$\mathbf{T} = \left[\begin{array}{ccc:c} & \mathbf{R} & & \mathbf{d} \\ \hdashline 0 & 0 & 0 & 1 \end{array}\right] \tag{2.4b}$$

However, this mapping is not exact because the reconstruction of the marker coordinates in the GCS is affected by measurement errors associated with skin movement, muscle contraction and stereophotogrammetric noise (Cappello *et al.*,

1997). The measurement errors can be minimised and an optimal solution for **R** and **d** can be determined by using the least-squares problem:

$$\min \sum_{i=1}^{9} \left\| \mathbf{R}f_i + \mathbf{d} - g_i \right\|^2 \tag{2.5}$$

Using the singular value decomposition (SVD) method, a solution can be found to (2.5) using the following steps:

1. Calculate $\bar{f}$ and $\bar{g}$ as shown in (2.6a) and (2.6b)

$$\bar{f} = \frac{1}{9}\sum_{i=1}^{9} f_i \tag{2.6a}$$

$$\bar{g} = \frac{1}{9}\sum_{i=1}^{9} g_i \tag{2.6b}$$

2. Calculate the vectors A and B using (2.7a) and (2.7b)

$$\mathbf{A} = \left[f_1 - \bar{f}, \ldots, f_9 - \bar{f}\right] \tag{2.7a}$$

$$\mathbf{B} = \left[g_1 - \bar{g}, \ldots, g_9 - \bar{g}\right] \tag{2.7b}$$

3. Calculate the matrix **D** using (2.8)

$$\mathbf{D} = \mathbf{B}\mathbf{A}^{\mathrm{T}} \tag{2.8}$$

where $\mathbf{A}^{\mathrm{T}}$ is the transpose of **A**

4. Compute the SVD of **D**:

$$\mathbf{P}\Gamma\mathbf{Q}^{\mathrm{T}} = \mathbf{D} \tag{2.9a}$$

where **P** and **Q** are 3×3 orthogonal matrices and

$$\mathbf{\Gamma} = \begin{pmatrix} \varphi_1 & 0 & 0 \\ 0 & \varphi_2 & 0 \\ 0 & 0 & \varphi_3 \end{pmatrix} \tag{2.9b}$$

where φ_1, φ_2, and φ_3 are the singular values of the matrix **D**.

5. Calculate the rotation matrix **R** using (2.10)

$$\mathbf{R} = \mathbf{PSQ}^{\mathrm{T}} \tag{2.10a}$$

where $\mathbf{Q}^{\mathrm{T}}$ is the transpose of **Q**, and

$$\mathbf{S} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \det\left(\mathbf{PQ}^{\mathrm{T}}\right) \end{pmatrix} \tag{2.10b}$$

6. Calculate the translation vector **d** using (2.11)

$$\mathbf{d} = \bar{g} - \mathbf{R}\bar{f} \tag{2.11}$$

The transformation matrix **T** is then defined using (2.4b).

$\mathbf{T}_{pg}$, $\mathbf{T}_{mg}$ and $\mathbf{T}_{ma}$ are all defined using this procedure with **f** and **g** defined as follows:

1. When calculating the transformation matrix $\mathbf{T}_{pg}$ that maps the PLCS coordinates of the three positions on the pointer, defined by **f**, to their GCS coordinates, defined by **g**, then $\mathbf{f} = \{100010001\}$ and $\mathbf{g} = \{origin_{p,g} + \mathbf{u}_4,\ origin_{p,g} + \mathbf{u}_1,\ origin_{p,g} + \mathbf{u}_3\}$, where $origin_{p,g}$ are the GCS coordinates of the PLCS origin and $\mathbf{u}_4$, $\mathbf{u}_1$ and $\mathbf{u}_3$ are defined in section 2.2.1.1.

2. When calculating the transformation matrix $\mathbf{T}_{mg}$ that maps the MLCS coordinates of the three positions on the marker cluster defined by **f**, to their GCS coordinates defined by **g**, then $\mathbf{f} = \{100010001\}$ and $\mathbf{g} = \{origin_{m,g} + \mathbf{v}_4,\ origin_{m,g} + \mathbf{v}_1,\ origin_{m,g} + \mathbf{v}_3\}$, where $origin_{m,g}$ are the GCS coordinates of the MLCS origin and $\mathbf{v}_4$, $\mathbf{v}_1$ and $\mathbf{v}_3$ are defined in section 2.2.1.2.

3. When calculating the transformation matrix $\mathbf{T}_{ma}$ that maps the MLCS coordinates of the three positions on the tibia defined by **f**, to their tibial ALCS coordinates defined by **g**, then $\mathbf{f} = \{origin_{a,m} + \mathbf{w}_1,\ origin_{a,m} + \mathbf{w}_3,\ origin_{a,m} + \mathbf{w}_4\}$ and $\mathbf{g} = \{100010001\}$, where $origin_{a,m}$ are the MLCS coordinates of the tibial ALCS origin and $\mathbf{w}_1$, $\mathbf{w}_3$ and $\mathbf{w}_4$ are defined in section 2.2.1.3.

4. When calculating the transformation matrix $\mathbf{T}_{ma}$ that maps the MLCS coordinates of the three positions on the femur defined by **f**, to their femoral ALCS coordinates defined by **g**, then $\mathbf{f} = \{origin_{a,m} + \mathbf{n}_1,\ origin_{a,m} + \mathbf{n}_3,\ origin_{a,m} + \mathbf{n}_4\}$ and $\mathbf{g} = \{100010001\}$, where $origin_{a,m}$ are the MLCS coordinates of the femoral ALCS origin and $\mathbf{n}_1$, $\mathbf{n}_3$ and $\mathbf{n}_4$ are defined in section 2.2.1.4.

2.2.3 The Joint Coordinate System (JCS) Approach

The JCS approach (Grood and Suntay, 1983) gives a simple geometric description of the 6-degree of freedom movement of the tibia relative to the femur in terms of clinical reference planes.

A JCS, composed of three axes, is established for the knee as shown in Figure 2.12. Two of the axes are embedded in the femur and tibia and as such are called fixed body axes. The tibial fixed body axis is its mechanical axis (the *z*-axis) as shown in Figure 2.11a. The fixed body axis in the femur corresponds to the *X*-axis as shown in Figure 2.11b. The third axis is perpendicular to both fixed body axes. Since this third axis is not fixed to either body, it is known as the floating axis.

The joint rotations and translations occur about these three joint coordinate axes as shown in Figure 2.12. Flexion-extension occurs about the femoral fixed body axis, external-internal rotation about the tibial fixed axis and abduction-adduction about the floating axis. Medial-lateral tibial shift occurs along the femoral axis, anterior-posterior tibial drawer along the floating axis and joint compression-distraction along the tibial fixed axis.

The knee joint rotations and translations can be determined from the 4 × 4 transformation matrix $\mathbf{T}_{tf}$. In this study, the joint translations were calculated, but only the knee joint rotations were considered since these constitute kinematic variables required for the present study and there are problems associated with the exact meaning of the translations in terms of their relationship to the flexion-extension of the knee. For the calculation of the joint translations, the reader is directed to Grood and Suntay (1983), and Lafortune *et al.*, (1992). If $\mathbf{T}_{tf}$ is defined according to the convention of (2.3), and if $T_{r,c}$ represents the element of $\mathbf{T}_{tf}$ in row r and column c then the rotations are calculated using equations (2.12) to (2.14) (Grood and Suntay, 1983):

$$\theta_{FE} = \tan^{-1}\left(\frac{T_{3,2}}{T_{3,3}}\right) \tag{2.12}$$

where θ_{FE} is the flexion-extension angle and flexion is defined as a positive angle (see Figure 2.12) An example of the flexion-extension waveform is given in Figure 2.13.

$$\theta_{AA} = \frac{\pi}{2} + \cos^{-1}\left(T_{3,1}\right) \tag{2.13a}$$

for the right knee where θ_{AA} is the adduction-abduction angle and abduction is defined as a positive angle (see Figure 2.12), or

$$\theta_{AA} = \cos^{-1}\left(T_{3,1}\right) - \frac{\pi}{2} \tag{2.13b}$$

for the left knee. An example of the abduction-adduction waveform is given in Figure 2.14.

$$\theta_{IE} = -\tan^{-1}\left(\frac{T_{2,1}}{T_{1,1}}\right) \tag{2.14}$$

where θ_{IE} is the internal-external rotation angle and external rotation is defined as a positive angle (see Figure 2.12). An example of the internal-external rotation waveform is given in Figure 2.15.

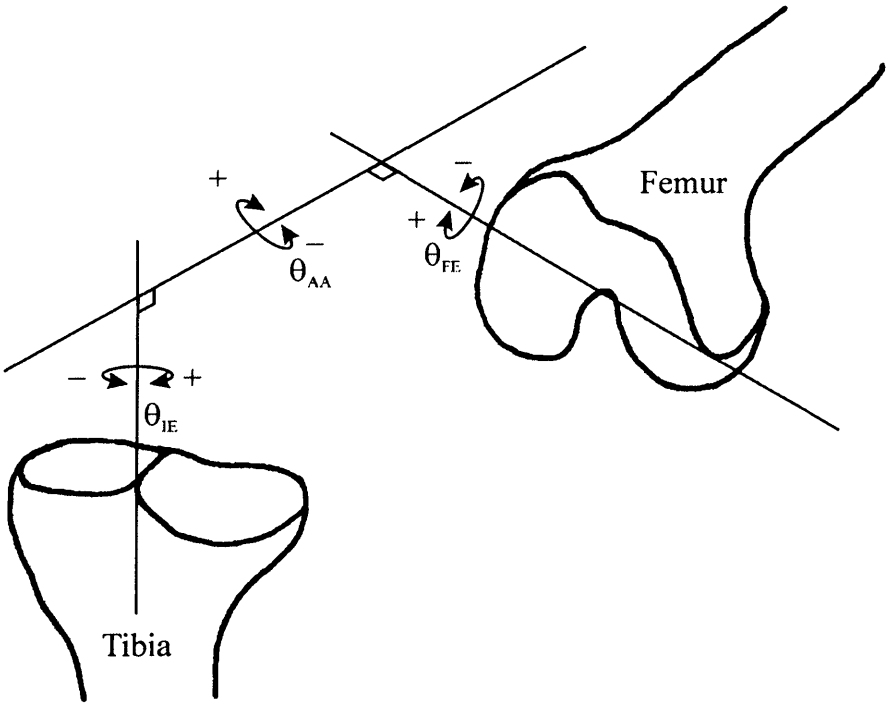


Figure 2.12 The joint coordinate system for the knee consists of two fixed body axes and a floating axis. The joint rotations (θ_{FE}, θ_{AA} and θ_{IE}) occur about these three axes.

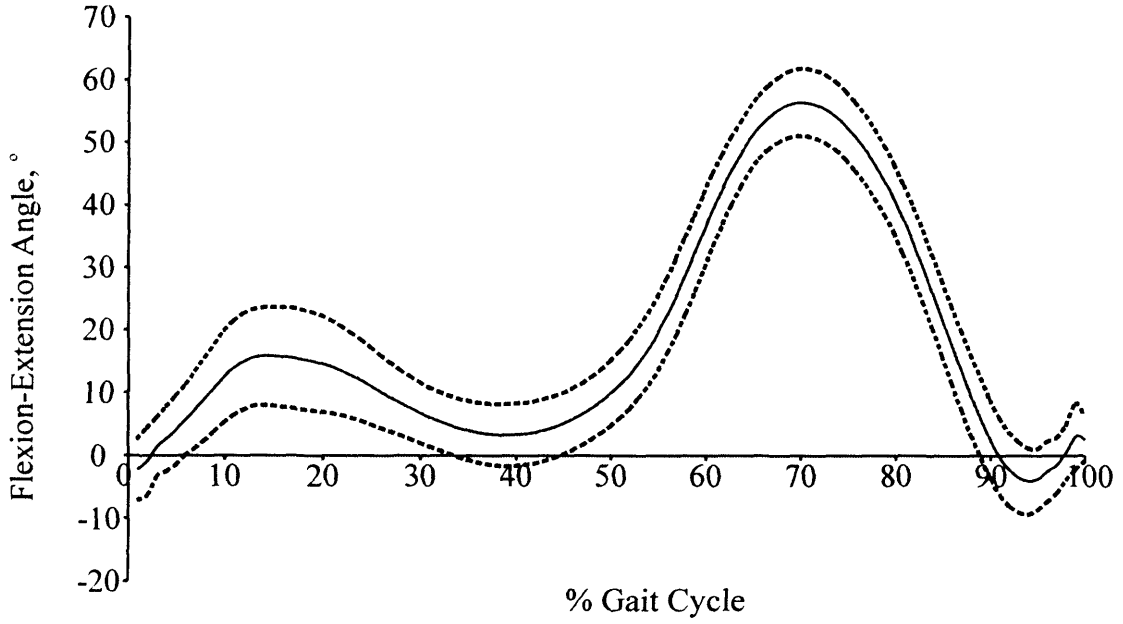


Figure 2.13 Mean flexion-extension waveform for a group of NL subjects (solid line). The dashed line indicates ± one standard deviation.

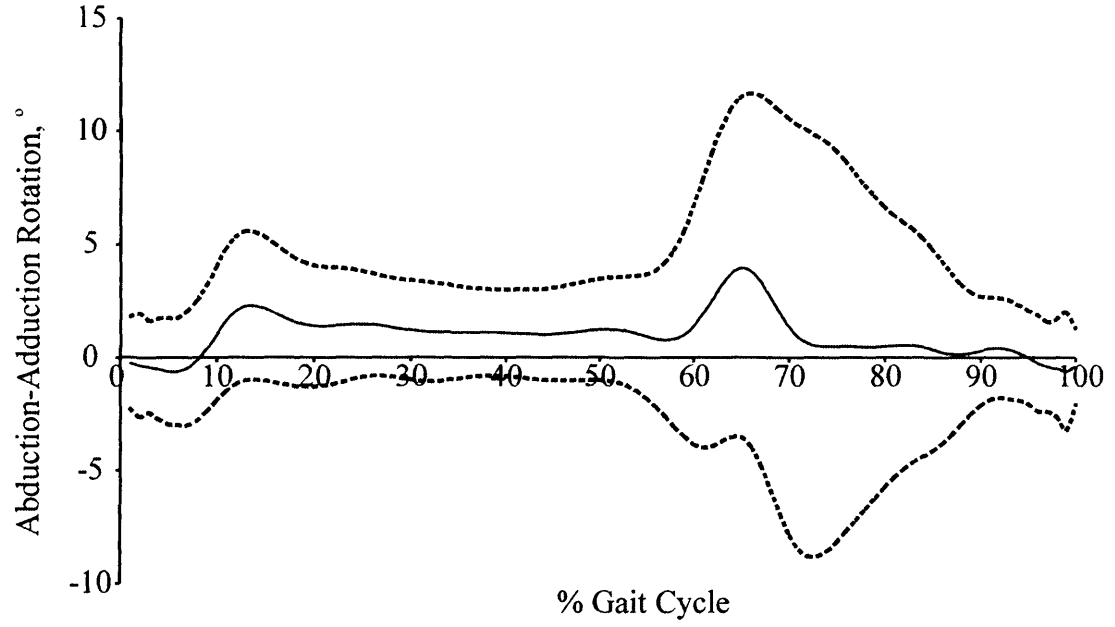


Figure 2.14 Mean abduction-adduction waveform for a group of NL subjects (solid line). The dashed line indicates ± one standard deviation.

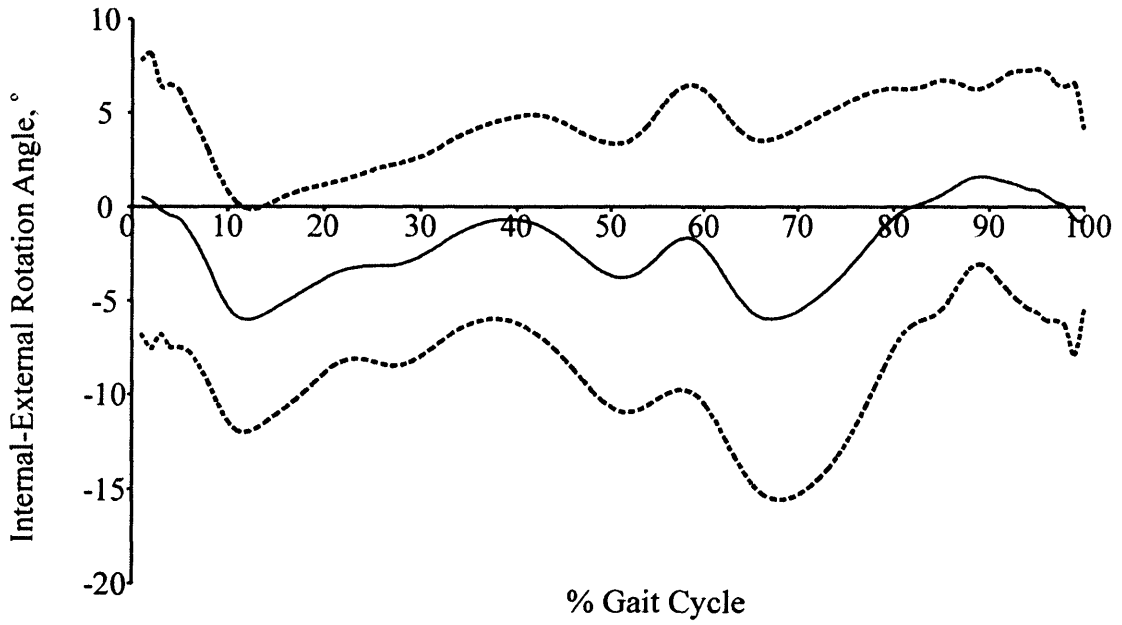


Figure 2.15 Mean internal external rotation waveform for a group of NL subjects (solid line). The dashed line indicates ± one standard deviation.

2.3 DATA PROCESSING - KINETICS

Previously developed Matlab (The MathWorks Inc.) software (Holt *et al.*, 2000) was applied to the raw signals produced by the force platform to calculate the three-dimensional ground reaction forces (GRFs) for the OA, NL and TKR samples.

Three force and three moment components are calculated using

$$\begin{bmatrix} F_x \\ F_y \\ F_z \\ M_x \\ M_y \\ M_z \end{bmatrix} = [\mathrm{D_F}] \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \end{bmatrix} \qquad (2.15a)$$

where F_x, F_y, F_z, M_x, M_y *and* M_z are the force and moment components in the force platform coordinate system as shown in Figure 2.16; S_1 to S_6 are the output signals divided by the amplifier gain; and $[\mathrm{D_F}]$ is the calibration matrix:

$$[D_F]=\begin{bmatrix} 1740.4 & -5.0 & -10.8 & 18.3 & 7.6 & 0 \\ 19.1 & 1737.3 & -9.8 & -9.7 & -12.6 & -10.5 \\ 40.4 & -40.9 & 3675.3 & 35.9 & -33.8 & -3.4 \\ -2.4 & -106.0 & -2.8 & 1143.7 & 5.1 & 2.6 \\ 112.2 & -0.8 & 2.1 & 2.4 & 804.1 & 2.8 \\ 1.5 & -2.4 & -7.0 & 1.7 & 1.9 & 401.5 \end{bmatrix} \quad (2.15b)$$

The medial-lateral (ML) GRF (F_{ML}), the anterior-posterior (AP) GRF (F_{AP}) and the vertical (V) GRF (F_V) are calculated from F_x, F_y and F_z as shown in (2.16) to (2.18)

$$F_{ML} = -F_x \quad (2.16)$$

where an anterior GRF is defined as a positive force.

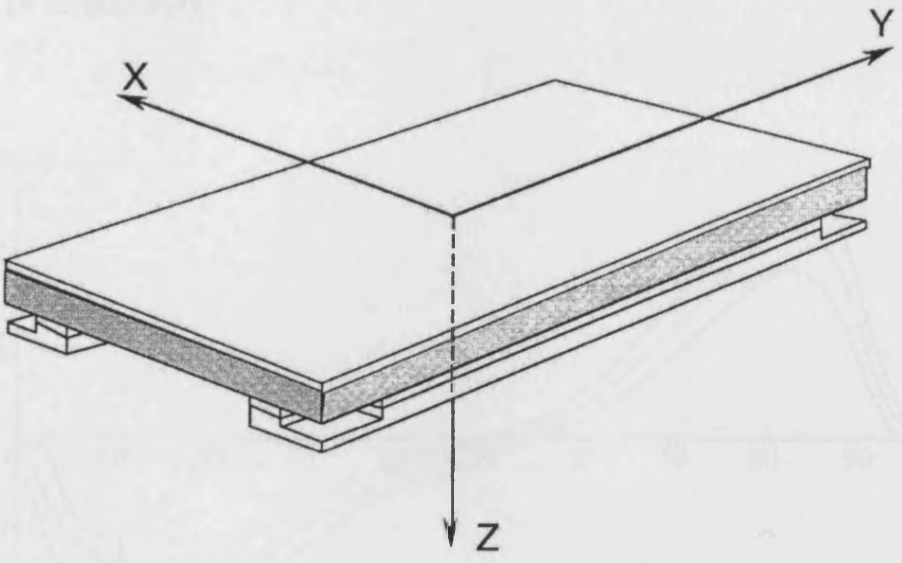


Figure 2.16 The force platform coordinate system (Taken from the User's Manual, Bertec Corporation).

For the right leg

$$F_{AP} = F_y \quad (2.17a)$$

and for the left leg

$$F_{AP} = -F_y \tag{2.17b}$$

where a lateral GRF is defined as a positive force.

$$F_V = F_z \tag{2.18}$$

The three GRF waveforms calculated using (2.16) to (2.18) were re-sampled at 60Hz (to match the sample rate of the kinematic waveforms) and subsequently re-sampled over 100% stance phase. The three GRF waveforms were normalised to body weight and an average of the six trials was computed. For some subjects, the ML GRF data was incomplete. Consequently, only the AP and V GRF waveforms are considered in this study. Examples of the AP GRF and the V GRF waveforms are given in Figures 2.17 and 2.18 respectively.

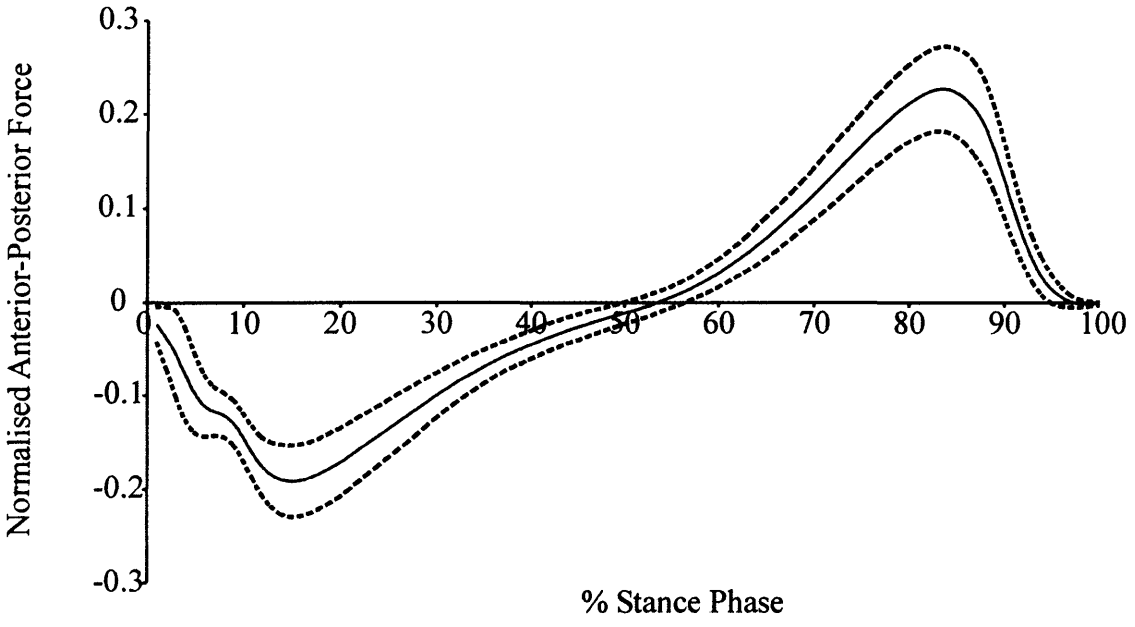


Figure 2.17 Mean AP GRF waveform normalised to body weight for a group of NL subjects (solid line). The dashed line indicates ± one standard deviation.

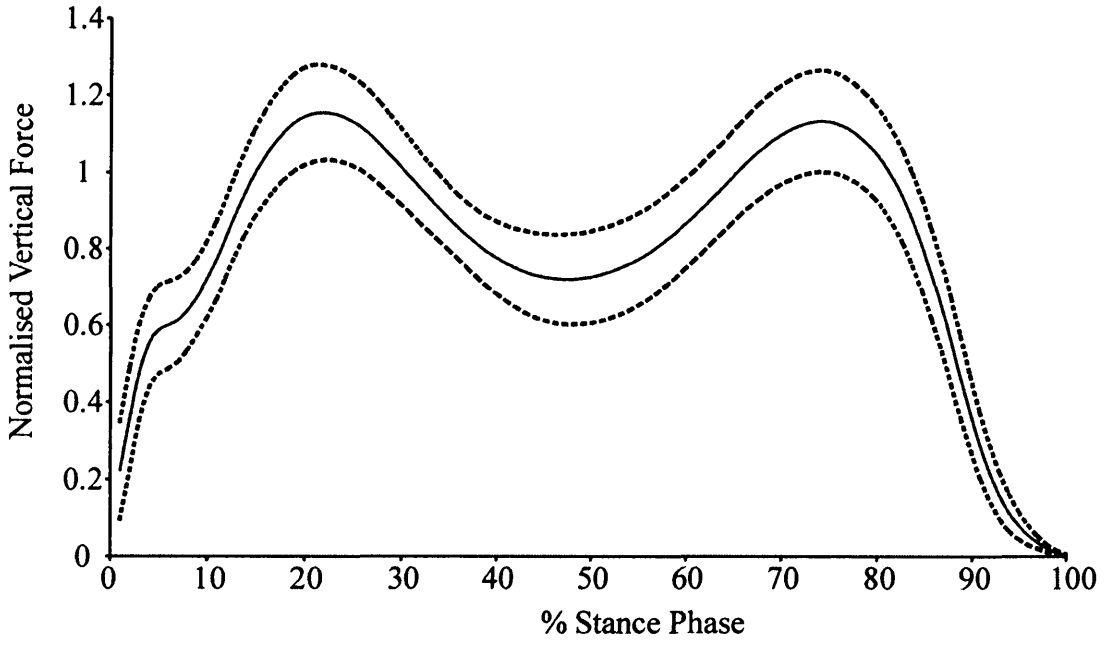


Figure 2.18 Mean V GRF waveform normalised to body weight for a group of NL subjects (solid line). The dashed line indicates ± one standard deviation.

2.4 DATA PROCESSING - ANTHROPOMETRICAL AND TEMPORAL DISTANCE PARAMETERS

Matlab (The MathWorks Inc.) software was developed to calculate a subject's body mass index (BMI), cadence and percentage stance phase for the OA, NL and TKR samples. BMI is defined as

$$\text{BMI} = \frac{w}{h^2} \tag{2.19}$$

where w and h were defined in section 2.1.2.

Although cadence is usually defined as the number of steps per minute (Whittle, 1996) it is here defined as the number of strides per minute, where a stride length is measured between successive placements of the same foot (Whittle, 1996).

2.5 DATA PROCESSING - REPRESENTATION OF TEMPORAL WAVEFORMS USING PRINCIPAL COMPONENT ANALYSIS

The kinematic and kinetic variables exist in the form of temporal waveforms as shown in Figures 2.13, 2.14, 2.15 2.17 and 2.18. As mentioned in Chapter 1, parameterisation of these waveforms is a common practice in gait analysis studies. However, a danger of this practice is that valuable temporal information is readily discarded. Following the work of Deluzio *et al.* (1997, 1999) as described in Chapter 1, Principal Component Analysis (PCA) was utilised to represent the gait waveforms in a discrete form whilst retaining temporal information. The general procedure for obtaining a PCA when using temporal waveforms is described and then applied specifically to the kinematic and kinetic waveforms.

2.5.1 PCA

A PCA is performed in a number of stages (Chau, 2001a; Tabachnick and Fidell, 1989; Daultrey, 1976):

1. Standardisation of data
2. Calculation of the correlation matrix
3. Eigendecomposition of the correlation matrix
4. Retention of principal components (PCs)
5. Calculation of the component loadings matrix
6. Calculation of the PC scores

These individual stages will now be discussed in turn.

2.5.1.1 Standardisation of data

The dataset consists of a time-normalised waveform sampled at each 1% from 0 to 100% of the gait cycle for a set of n subjects. Here each 1% of the waveform is referred to as a variable. Let the values of these variables for the i^{th} subject in the sample be y_{ij}, (j = 1: 100) respectively.

The first stage of a PCA is to standardise the entire dataset so that each variable has zero mean and unit standard deviation (Chau, 2001a). The standardised variables, z_{ij} are given as:

$$z_{ij} = \frac{y_{ij} - \bar{y}_j}{\sigma_j} \tag{2.20}$$

where z_{ij} is the standardised variable and $\bar{y}_j$ and σ_j are the sample mean and standard deviation for variable y_j respectively.

2.5.1.2 Calculation of the correlation matrix

The next stage of a PCA is to calculate the correlation matrix, **C**, for the entire set of standardised variables using (2.21) (Chau, 2001a):

$$\mathbf{C} = \frac{\mathbf{Z}^{\mathrm{T}}\mathbf{Z}}{(p-1)} = \begin{bmatrix} 1 & c_{12} & \cdots & c_{1,100} \\ c_{21} & 1 & \cdots & c_{2,100} \\ \vdots & \vdots & & \vdots \\ c_{100,1} & c_{100,2} & \cdots & 1 \end{bmatrix} \tag{2.21}$$

where p is the number of input variables, **Z** is the $n \times 100$ matrix containing the standardised variables z_j (j = 1:100), $\mathbf{Z}^{\mathrm{T}}$ is its transpose and $c_{kl} = c_{lk}$ is the correlation coefficient between variables z_k and z_l respectively.

2.5.1.3 Eigendecomposition of the correlation matrix

The subsequent step is to find the eigendecomposition of **C** using (2.21) (Daultrey, 1976; Tabachnick and Fidell, 1989; Chau, 2001a):

$$\mathbf{C} = \mathbf{E}\Lambda\mathbf{E}^{\mathrm{T}} \tag{2.22}$$

where Λ is the diagonal matrix of eigenvalues of **C**, λ_j (j = 1: 100), and the columns of the matrix **E** are the corresponding eigenvectors, e_j (j = 1: 100). The variance of the j^{th}

PC is given by λ_j. The first PC has the largest associated variance whilst the last PC has the smallest variance, i.e. $\lambda_1 > \lambda_2 > \ldots > \lambda_{100}$.

2.5.1.4 Retention of PCs

Although PCA produces the same number of PCs as there are original variables, it is hoped that the variances of the majority of PCs will be negligible and that the original variables can be described by a smaller number of PCs, *m*. Jolliffe (1986) reports numerous possible methods for determining the number of PCs needed to adequately explain the original data. Among these methods are two that are reported most often in the literature in the context of gait analysis (Yamamoto *et al.*, 1983; Sadeghi *et al.*, 1997; Sadeghi *et al.*, 2000).

The first method, Kaiser's rule, selects the *m* PCs by examining the size of their individual variances. The method is based on the premise that any PC with a variance less than one contains less information than the original variables (which have unit variance) and is therefore not worth retaining. The second method selects the *m* PCs by examining the cumulative percentage of total variation that they explain. The total variance, t_m, accounted for by the first *m* PCs is given by Joliffe (1986) as

$$t_m = \frac{100}{p} \sum_{j=1}^{m} \lambda_j \tag{2.23}$$

where *p* is the number of input variables. The number of PCs required to explain *q*% of the variation of the original data, is the smallest value of *m* for which $t_m \geq q$.

2.5.1.5 Calculation of the component loadings matrix

The penultimate stage of the PCA is to assign meaningful labels to each PC. This is accomplished through examination of the matrix of component loadings, **L**, which gives the weighted relationship between the PCs and the original variables. **L** is calculated using (2.24a) (Tabachnick and Fidell, 1989; Daultrey, 1976):

$$\mathbf{L} = \mathbf{E}\boldsymbol{\Lambda}^{\frac{1}{2}} = \begin{bmatrix} l_{1,1} & l_{1,2} & \cdots & l_{1,100} \\ l_{2,1} & l_{2,2} & \cdots & l_{2,100} \\ \vdots & \vdots & & \vdots \\ l_{100,1} & l_{100,2} & \cdots & l_{100,100} \end{bmatrix} \tag{2.24a}$$

where $l_{i,j}$ are the correlation coefficients between the i^{th} variable and the j^{th} PC and

$$l_{i1}^{\,2} + l_{i2}^{\,2} + l_{i3}^{\,2} + \ldots + l_{i100}^{\,2} = 1 \tag{2.24b}$$

To interpret and thus assign a label to a PC, a threshold value is selected (Comrey, 1973; Sadeghi *et al.*, 1997; Sadeghi *et al.*, 2000; Tabachnick and Fidell, 1989). The variables with loadings above this threshold are collated and "the researcher searches for a concept that unifies them" (Tabachnick and Fidell, *pp.*639). Comrey (1973) suggests using a threshold value of 0.71. If a variable, say z_1, has a component loading of 0.71 on one PC, say $l_{1,1}$, then from (2.24b)

$$\begin{aligned} l_{1,1}^{\,2} &= 1 - l_{1,2}^{\,2} - l_{1,3}^{\,2} - \ldots - l_{1,100}^{\,2} = 0.5041 \\ \Rightarrow \sum_{j=2}^{100} l_{1,j}^{\,2} &= 0.4959 = \sqrt{0.7042} \end{aligned} \tag{2.25}$$

Consequently, from (2.25) the maximum value that $l_{1,j}$ (j = 2:100) can take is 0.7042. Therefore, using a threshold value of 0.71 implies that a variable can only load against one component and subsequently each PC will have a different interpretation. As a result a threshold value of 0.71 is used in this study.

2.5.1.6 Calculation of the PC scores

The final stage of the PCA is to calculate PC scores, $\boldsymbol{\Omega}$ for each individual in the sample using (2.26) (Daultrey, 1976; Tabachnick and Fidell, 1989):

$$\boldsymbol{\Omega} = \mathbf{ZE} \tag{2.26}$$

where **Z** is the matrix containing the standardised variables z_j (j = 1:100) and the columns of the matrix **E** are eigenvectors, e_j (j = 1:100). Thus, the PC scores are linear combinations of the original variables.

2.5.2 PCA Results

Following the procedure described in sections 2.2.4.1 to 2.2.4.6, custom Matlab software was written to obtain PC representations of the anterior-posterior GRF (APF), vertical GRF (VF), flexion-extension rotation (FER), abduction-adduction rotation (AAR) and internal-external rotation (IER) waveforms. The PCA procedure was applied to the OA and NL samples. The application of PCA to the TKR sample is described in section 2.5.3. To aid in interpretation of the PCs a timing of gait events is depicted in Figure 2.19 and recorded in Table 2.2.

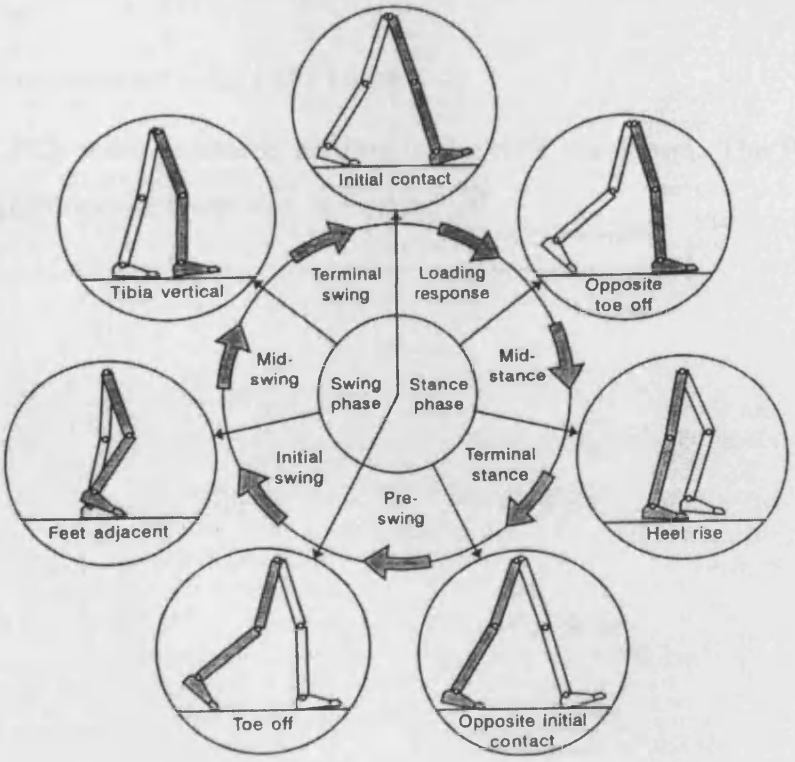


Figure 2.19 Phases and events of a single gait cycle of the right leg (Taken from Whittle, 1996, *pp.*59).

Table 2.2 Timing of gait events (adapted from Rose and Gamble, 1994, *pp.*143).

Gait cycle event	Timing of event (% gait cycle)	Timing of event (% stance phase)
Initial contact	Instantaneous	Instantaneous
Loading response	0 – 12	0 – 16
Mid-stance	12 – 30	16 – 48
Terminal stance	30 – 50	48 – 81
Pre-swing	50 – 62	81 – 100
Initial swing	62 – 75	–
Mid-swing	75 – 85	–
Terminal swing	85 – 100	–

2.5.2.1 Anterior-posterior GRF (APF) results

One hundred PCs were produced relating to the APF waveform. The PCs and their associated eigenvalues are depicted in Figure 2.20.

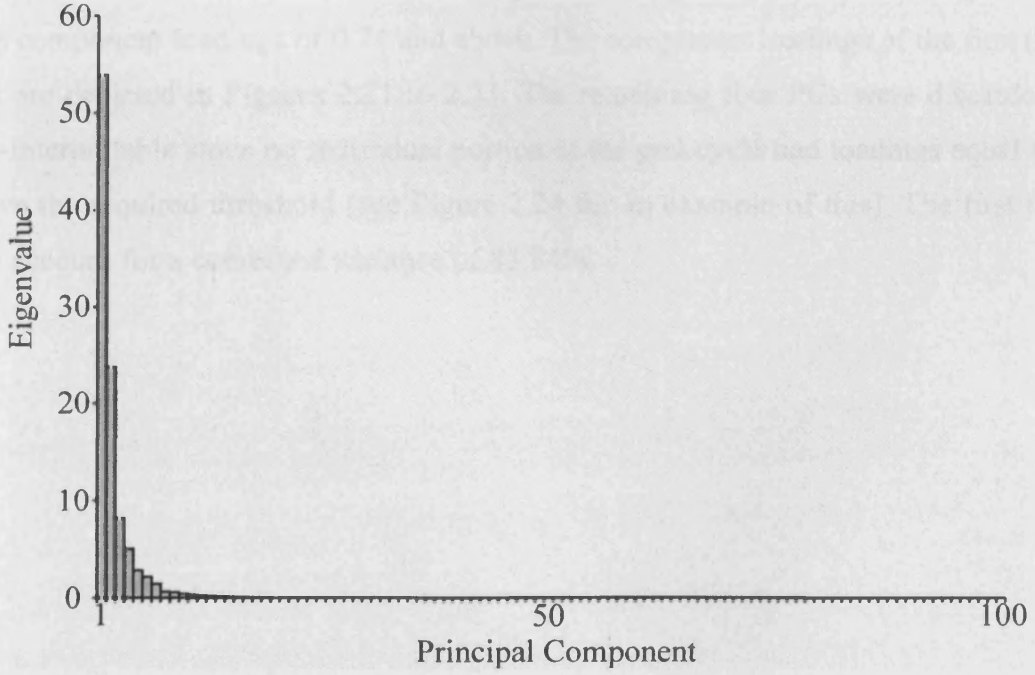


Figure 2.20 The eigenvalues of the 100 PCs for the APF waveform.

The first few PCs explain most of the variation in the original data. After the first few PCs, the values of the associated eigenvalues rapidly approach zero. Using Kaiser's rule, seven PCs were retained. The eigenvalues of these seven PCs are recorded in Table 2.3.

Table 2.3 Eigenvalues of the seven APFPCs retained using Kaiser's rule.

Principal Component	Eigenvalue
APFPC1	53.921
APFPC2	23.762
APFPC3	8.156
APFPC4	4.993
APFPC5	2.811
APFPC6	2.150
APFPC7	1.450

These seven PCs account for a combined variance of 97.24%. Following Comrey (1973), the seven PCs were interpreted by identifying the portions of the gait cycle with component loadings of 0.71 and above. The component loadings of the first three PCs are depicted in Figures 2.21 to 2.23. The remaining four PCs were discarded as non-interpretable since no individual portion of the gait cycle had loadings equal to or above the required threshold (see Figure 2.24 for an example of this). The first three PCs account for a combined variance of 85.84%.

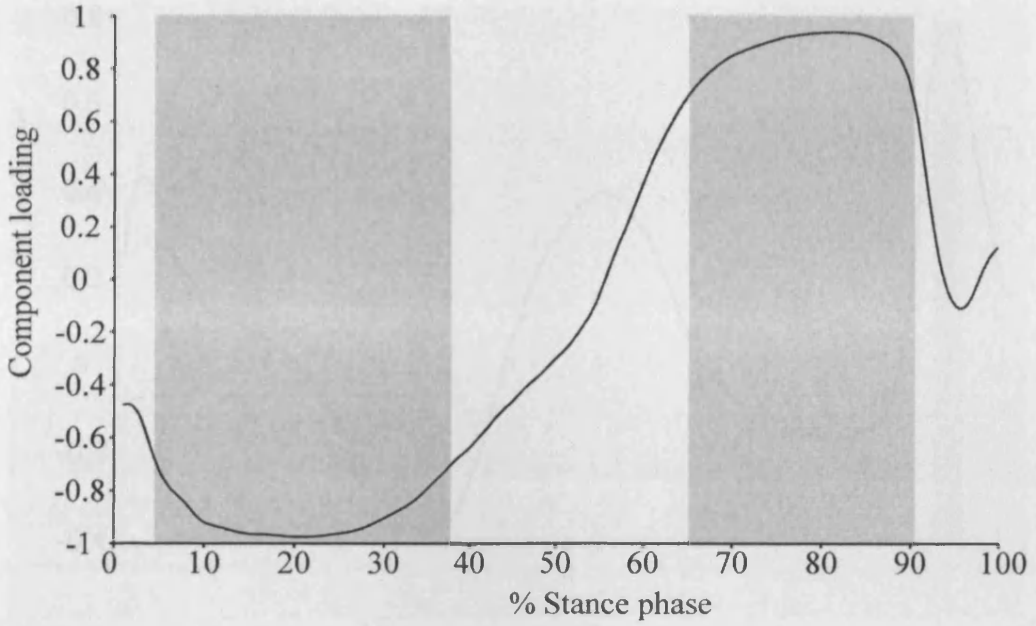


Figure 2.21 The component loadings of the first PC for the APF waveform (APFPC1). The grey shaded areas indicate the portions of the gait cycle with component loadings of 0.71 or greater.

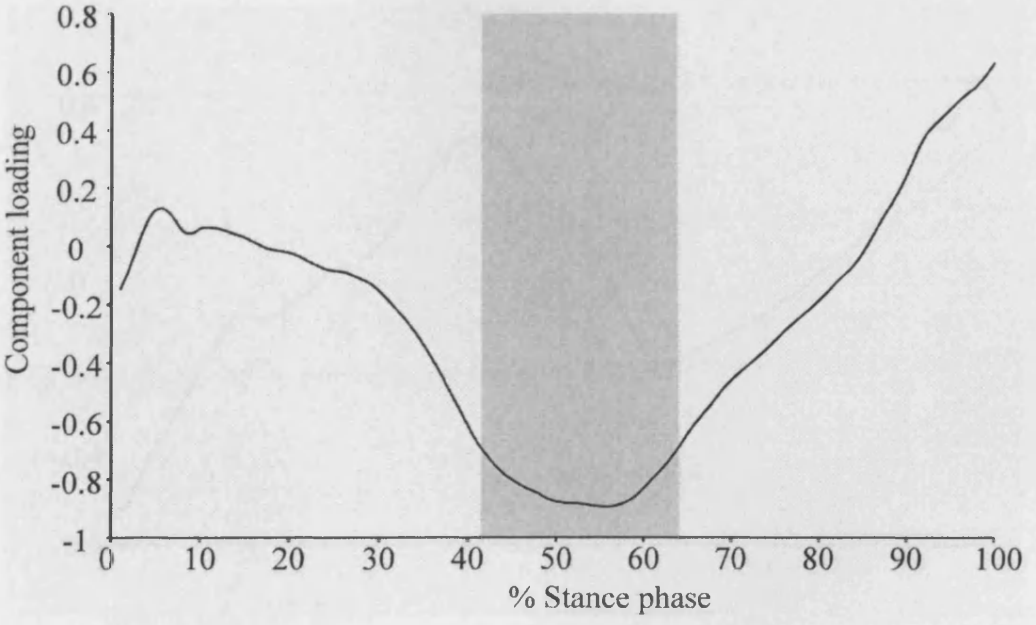


Figure 2.22 The component loadings of second PC for the APF waveform (APFPC2). The grey shaded area indicates the portions of the gait cycle with component loadings of 0.71 or greater.

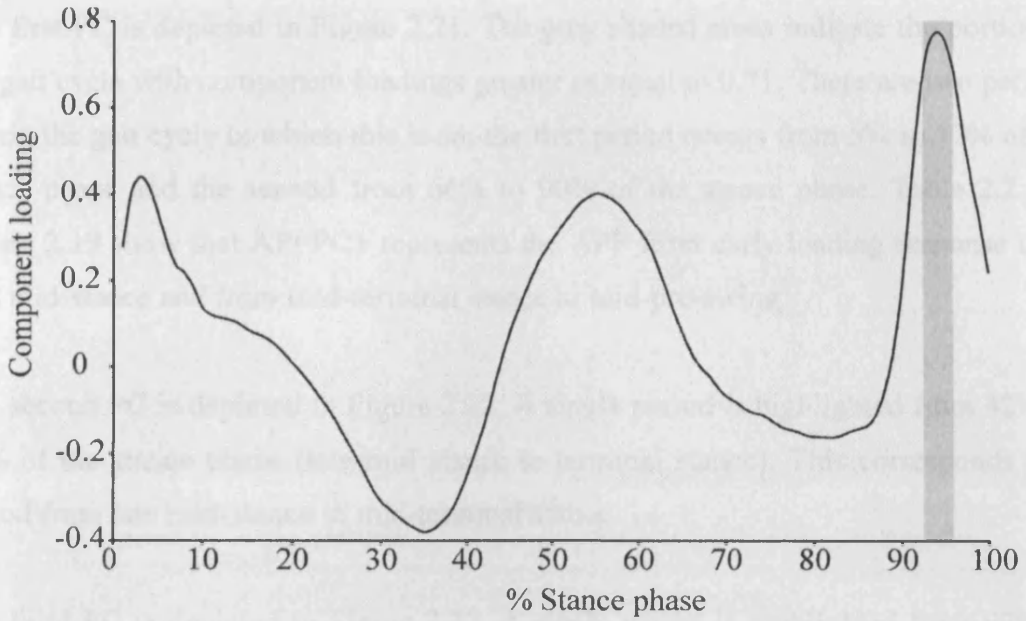


Figure 2.23 The component loadings of the third PC (APFPC3) for the APF waveform. The grey shaded area indicates the portions of the gait cycle with component loadings of 0.71 or greater.

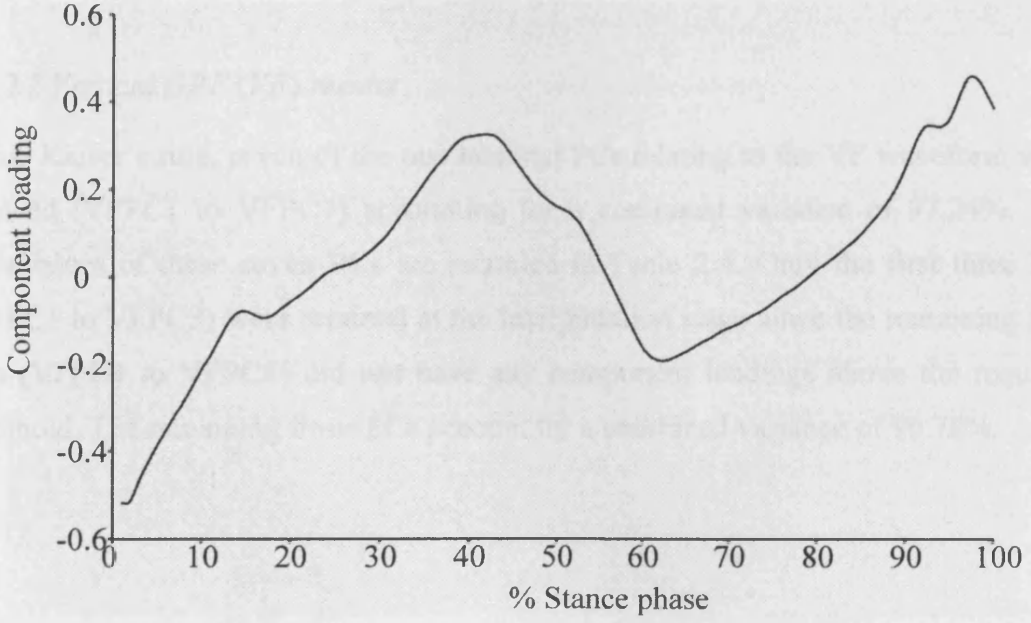


Figure 2.24 The component loadings of the fourth PC for the APF waveform (APFPC4). None of the portions of the gait cycle have component loadings of 0.71 or greater.

The first PC is depicted in Figure 2.21. The grey shaded areas indicate the portion of the gait cycle with component loadings greater or equal to 0.71. There are two periods within the gait cycle in which this is so: the first period occurs from 5% to 37% of the stance phase and the second from 66% to 90% of the stance phase. Table 2.2 and Figure 2.19 show that APFPC1 represents the APF from early loading response until mid mid-stance and from mid-terminal stance to mid-pre-swing.

The second PC is depicted in Figure 2.22. A single period is highlighted from 42% to 63% of the stance phase (late mid stance to terminal stance). This corresponds to a period from late mid-stance to mid-terminal stance.

The third PC is depicted in Figure 2.23. A single period is highlighted from 93% to 95% of the stance phase, a period during late pre-swing. It can be seen that each of the PCs has a different interpretation, each accounting for different portions of the stance phase. PC scores were calculated for the OA and NL samples for APFPC1, APFPC2 and APFPC3.

2.5.2.2 Vertical GRF (VF) results

Using Kaiser's rule, seven of the one hundred PCs relating to the VF waveform were retained (VFPC1 to VFPC7) accounting for a combined variance of 97.24%. The eigenvalues of these seven PCs are recorded in Table 2.4. Only the first three PCs (VFPC1 to VFPC3) were retained at the interpretation stage since the remaining four PCs (VFPC4 to VFPC7) did not have any component loadings above the required threshold. The remaining three PCs account for a combined variance of 90.78%.

Table 2.4 Eigenvalues of the seven VFPCs retained using Kaiser's rule.

Principal Component	Eigenvalue
VFPC1	43.508
VFPC2	34.271
VFPC3	12.996
VFPC4	3.315
VFPC5	2.135
VFPC6	1.473
VFPC7	1.020

The component loadings of VFPC1 are depicted in Figure 2.25. The grey shaded areas indicate the portion of the gait cycle with component loadings greater or equal to 0.71. Two periods are highlighted 28% to 42% (mid stance) and 49% to 81% (terminal stance). Table 2.2 and Figure 2.19 show that VFPC1 represents a portion of mid-stance and the portion from heel-rise to opposite initial contact.

The component loadings of VFPC2 are highlighted in Figure 2.26. This component is related to the ground reaction force in the period of 2 to 27% stance phase i.e. from loading response to mid-stance.

The component loadings of VFPC3 are shown in Figure 2.27 and a period from 86% to 100% stance phase is highlighted. VFPC3 represents the VF waveform during pre-swing. PC scores were subsequently calculated for the OA and NL samples for VFPC1, VFPC2 and VFPC3.

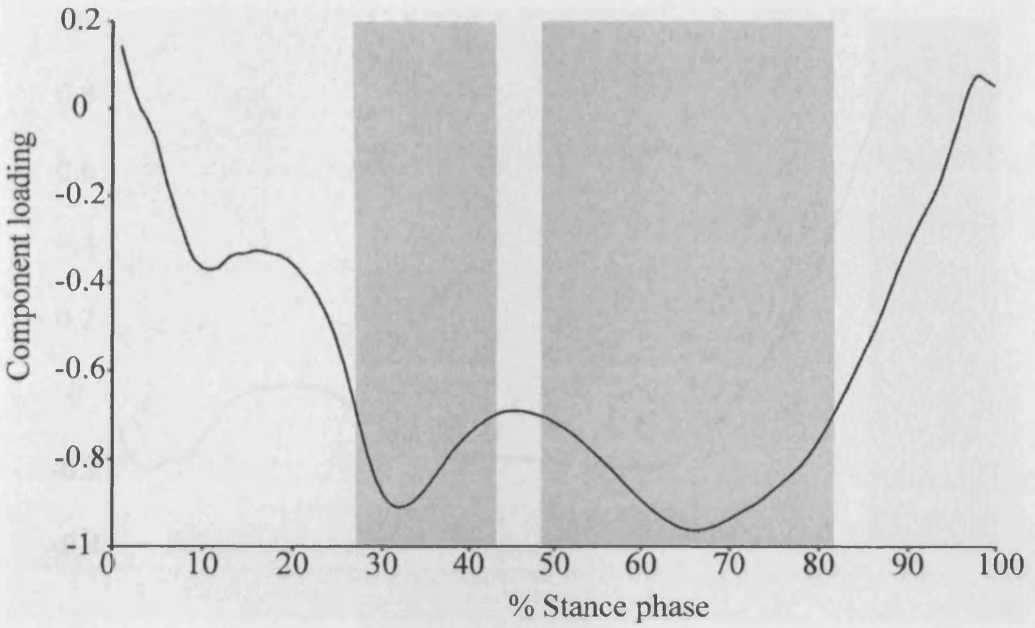


Figure 2.25 The component loadings of the first PC for the VF waveform (VFPC1). The grey shaded areas indicate the portions of the gait cycle with component loadings of 0.71 or greater.

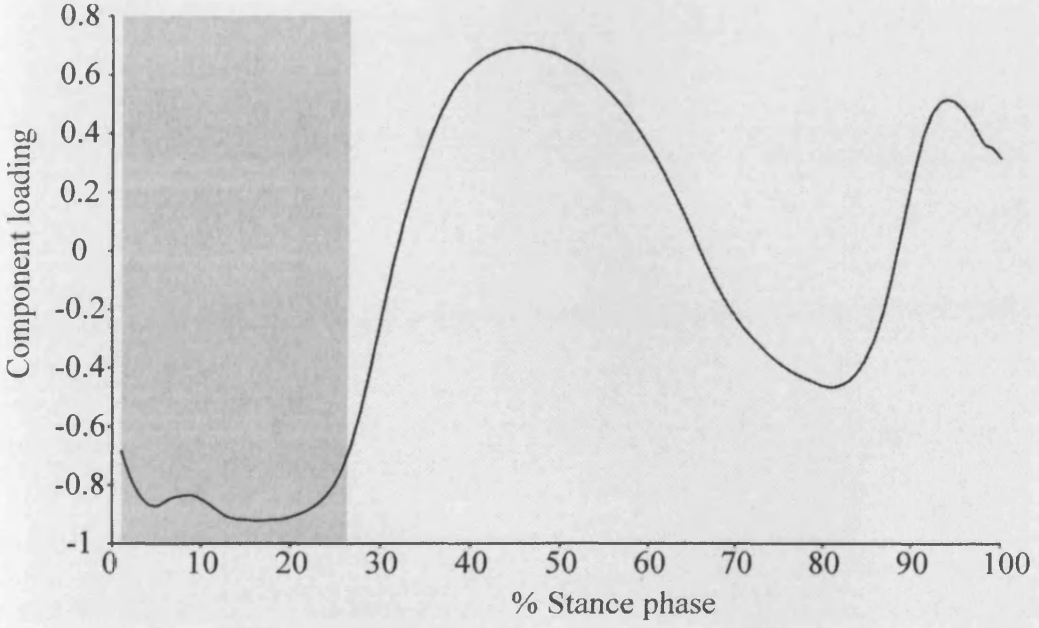


Figure 2.26 The component loadings of the second PC for the VF waveform (VFPC2). The grey shaded area indicates the portions of the gait cycle with component loadings of 0.71 or greater.

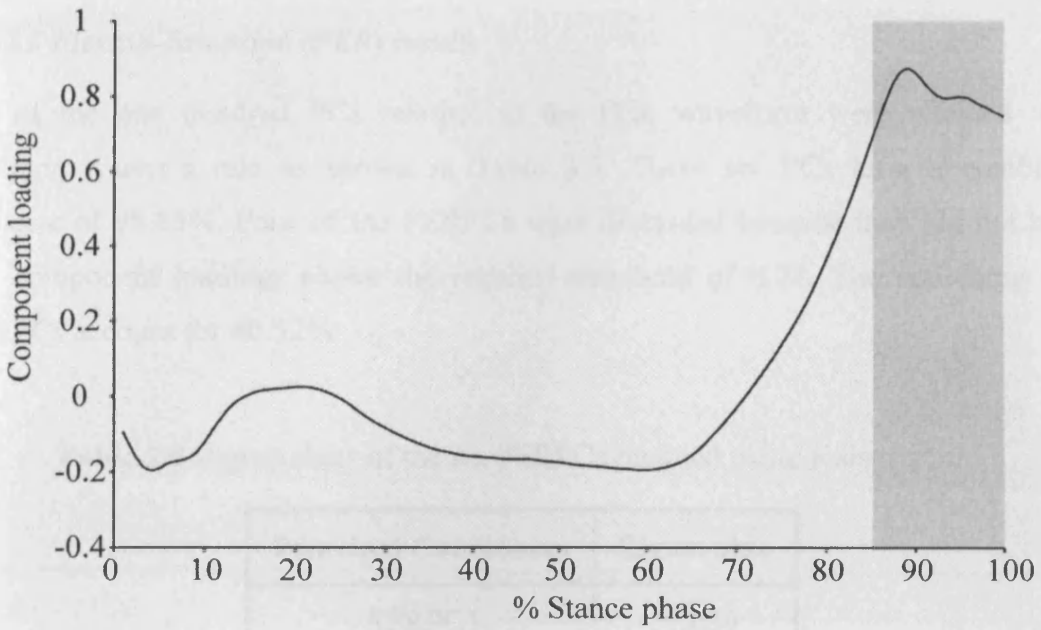


Figure 2.27 The component loadings of the third PC for the VF waveform (VFPC3). The grey shaded areas indicate the portions of the gait cycle with component loadings of 0.71 or greater.

2.5.2.3 Flexion-Extension (FER) results

Six of the one hundred PCs relating to the FER waveform were retained after applying Kaiser's rule as shown in Table 2.5. These six PCs have a combined variance of 98.85%. Four of the FERPCs were discarded because they did not have any component loadings above the required threshold of 0.71. The remaining two FERPCs account for 80.52%.

Table 2.5 Eigenvalues of the six FERPCs retained using Kaiser's rule.

Principal Component	Eigenvalue
FERPC1	57.039
FERPC2	23.484
FERPC3	9.830
FERPC4	5.313
FERPC5	1.634
FERPC6	1.551

The component loadings of these two FERPCs are shown in Figures 2.28 and 2.29 respectively. FERPC1 has component loadings above the threshold from 1 to 54% of the gait cycle as shown in Figure 2.28. Table 2.2 and Figure 2.19 show that FEPC1 represents the FER over a period from initial contact to opposite initial contact.

The component loadings of FEPC2 are depicted in Figure 2.29 where a period from 58 to 76% gait cycle is highlighted. PC scores were calculated for the OA and NL samples for FERPC1 and FERPC2.

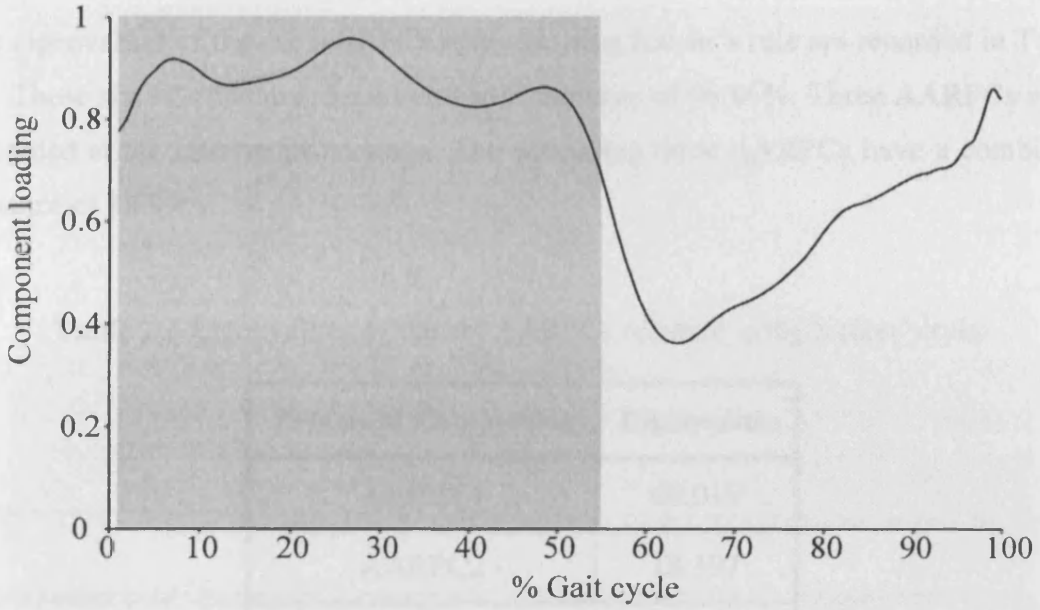


Figure 2.28 The component loadings of the first PC for the FER waveform (FERPC1). The grey shaded area indicates the portions of the gait cycle with component loadings of 0.71 or greater.

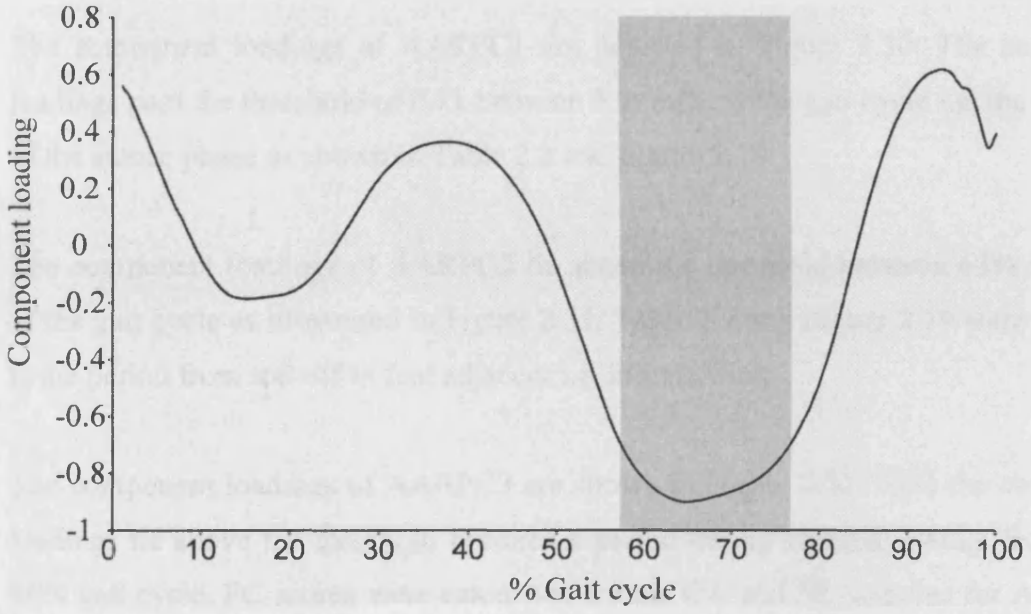


Figure 2.29 The component loadings of the second PC for the FER waveform (FERPC2). The grey shaded areas indicate the portions of the gait cycle with component loadings of 0.71 or greater.

2.5.2.4 Abduction-Adduction (AAR) results

The eigenvalues of the six AARPCs retained using Kaiser's rule are recorded in Table 2.6. These six PCs account for a combined variance of 96.86%. Three AARPCs were discarded at the interpretation stage. The remaining three AARPCs have a combined variance of 88.89%.

Table 2.6 Eigenvalues of the six AARPCs retained using Kaiser's rule.

Principal Component	Eigenvalue
AARPC1	60.019
AARPC2	18.397
AARPC3	10.473
AARPC4	5.193
AARPC5	1.675
AARPC6	1.104

The component loadings of AARPC1 are depicted in Figure 2.30. The component loadings pass the threshold of 0.71 between 3 to 60% of the gait cycle i.e. the majority of the stance phase as shown in Table 2.2 and Figure 2.19.

The component loadings of AARPC2 lie above the threshold between 63% and 75% of the gait cycle as illustrated in Figure 2.31. Table 2.2 and Figure 2.19 show that this is the period from toe-off to feet adjacent i.e. initial swing.

The component loadings of AARPC3 are shown in Figure 2.32. Here the component loadings lie above the threshold between a period during terminal swing: from 92 to 95% gait cycle. PC scores were calculated for the OA and NL samples for AARPC1, AARPC2 and AARPC3.

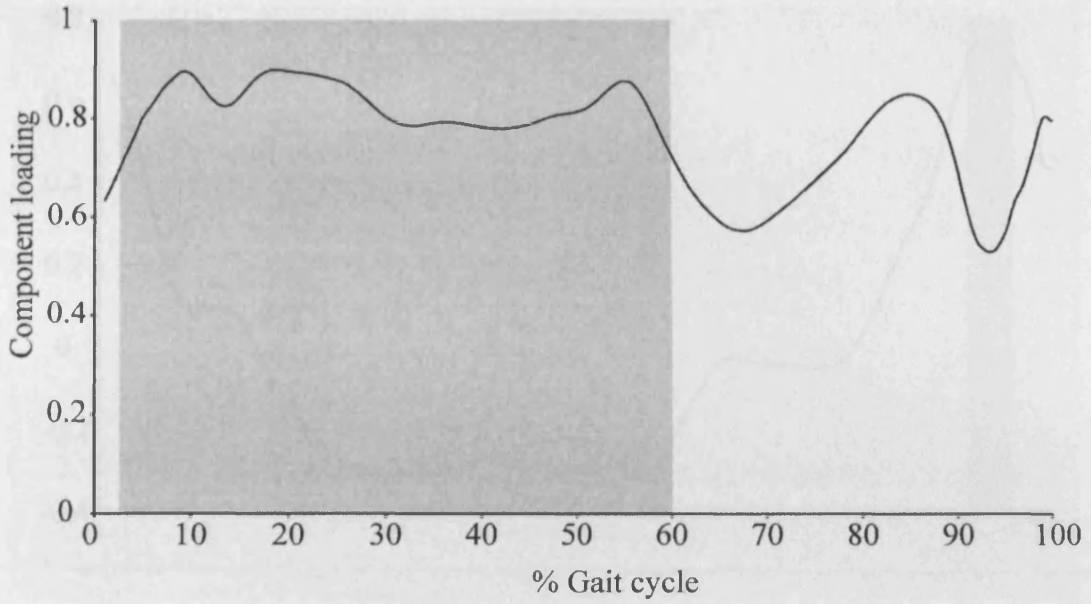


Figure 2.30 The component loadings of the first PC for the AAR waveform (AARPC1). The grey shaded area indicates the portions of the gait cycle with component loadings of 0.71 or greater.

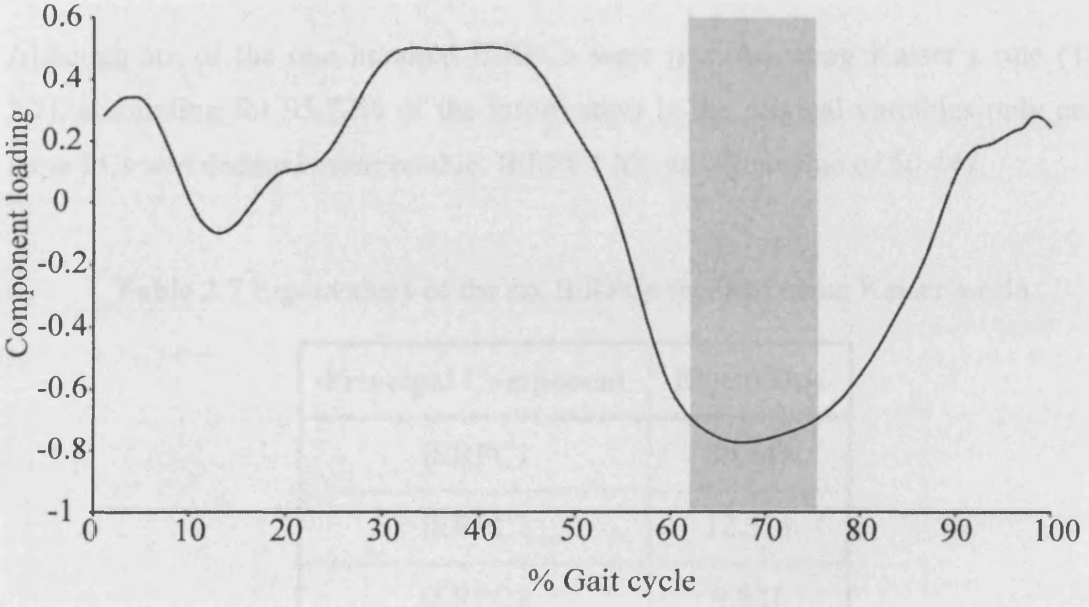


Figure 2.31 The component loadings of the second PC for the AAR waveform (AARPC2). The grey shaded areas indicate the portions of the gait cycle with component loadings of 0.71 or greater.

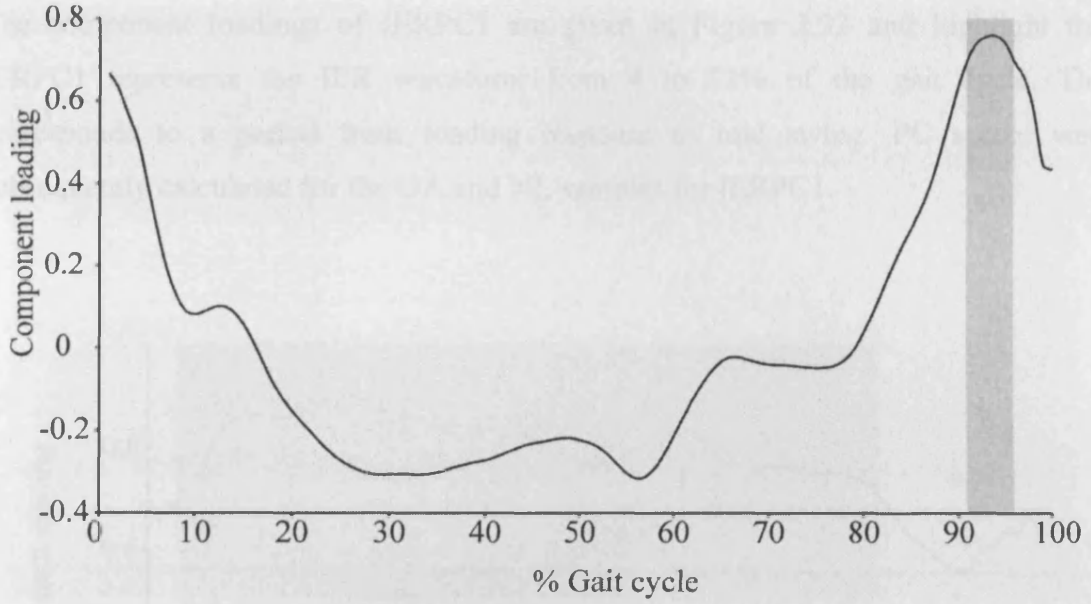


Figure 2.32 The component loadings of the third PC for the AAR waveform (AARPC3). The grey shaded areas indicate the portions of the gait cycle with component loadings of 0.71 or greater.

2.5.2.5 Internal-External Rotation (IER) results

Although six of the one hundred IERPCs were retained using Kaiser's rule (Table 2.7), accounting for 95.72% of the information in the original variables only one of these PCs was deemed interpretable. IERPC1 has an eigenvalue of 60.447.

Table 2.7 Eigenvalues of the six IERPCs retained using Kaiser's rule.

Principal Component	Eigenvalue
IERPC1	60.447
IERPC2	12.515
IERPC3	9.021
IERPC4	6.577
IERPC5	5.388
IERPC6	1.768

The component loadings of IERPC1 are given in Figure 2.33 and highlight that IERPC1 represents the IER waveform from 4 to 82% of the gait cycle. This corresponds to a period from loading response to mid swing. PC scores were subsequently calculated for the OA and NL samples for IERPC1.

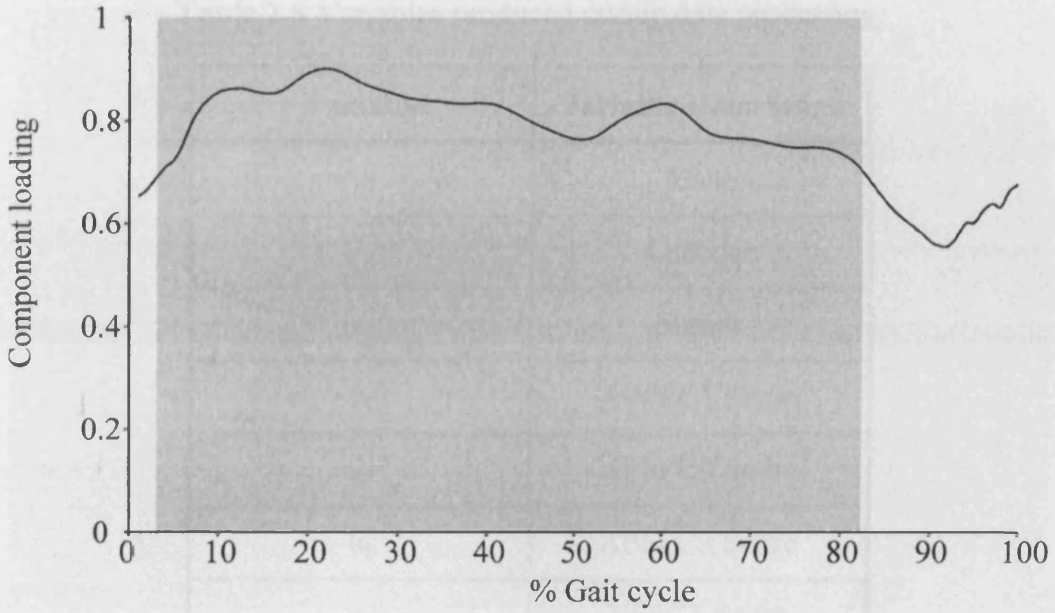


Figure 2.33 The component loadings of the first PC for the IER waveform (IERPC1). The grey shaded areas indicate the portions of the gait cycle with component loadings of 0.71 or greater.

The interpretation stage of this PCA is highly subjective. However, the pre-processing of the kinetic and kinematic waveforms using PCA means that important temporal information is retained, because the parametric representations relate to specific and unique portions of the gait cycle.

2.5.3 Application of PCA to TKR sample

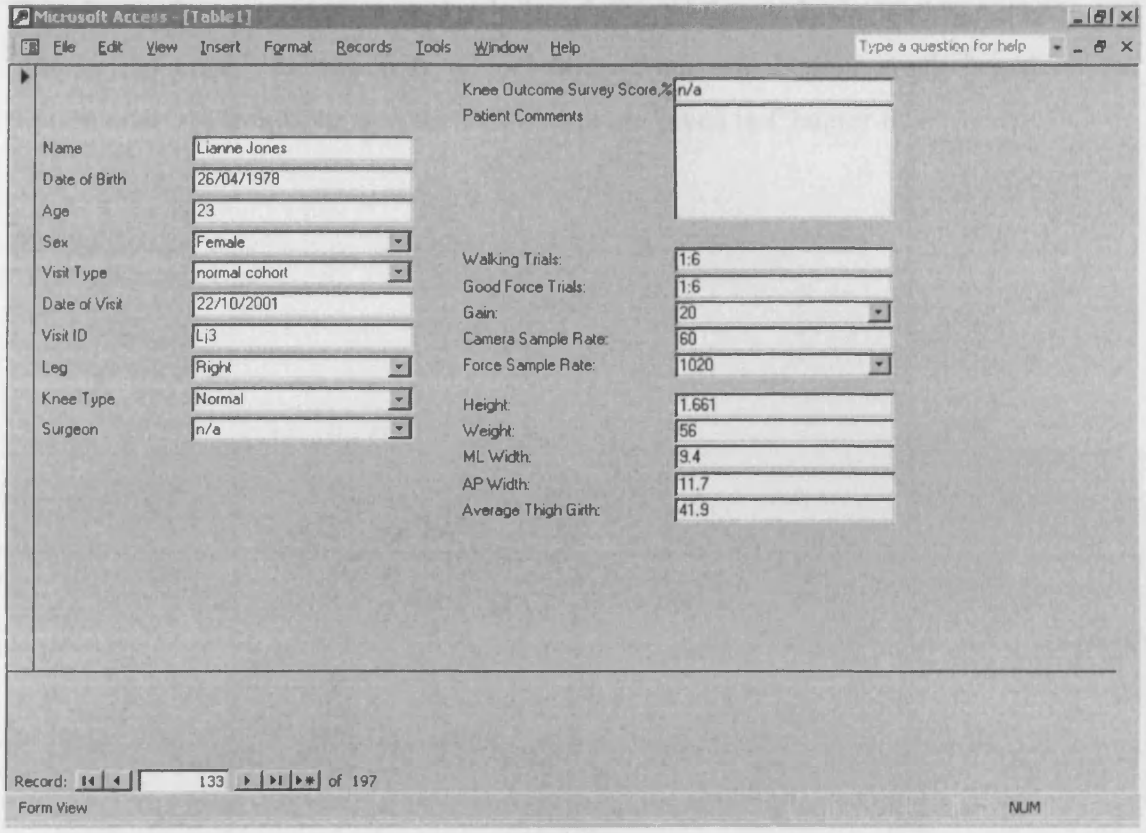
The variables in the TKR sample were normalised following the procedure in 2.5.11 using the mean and standard deviation of the combined OA and NL sample. The PC scores for the twelve PCs produced in section 2.5.2 were then calculated for the TKR sample using (2.26)

2.6 DATA STORAGE

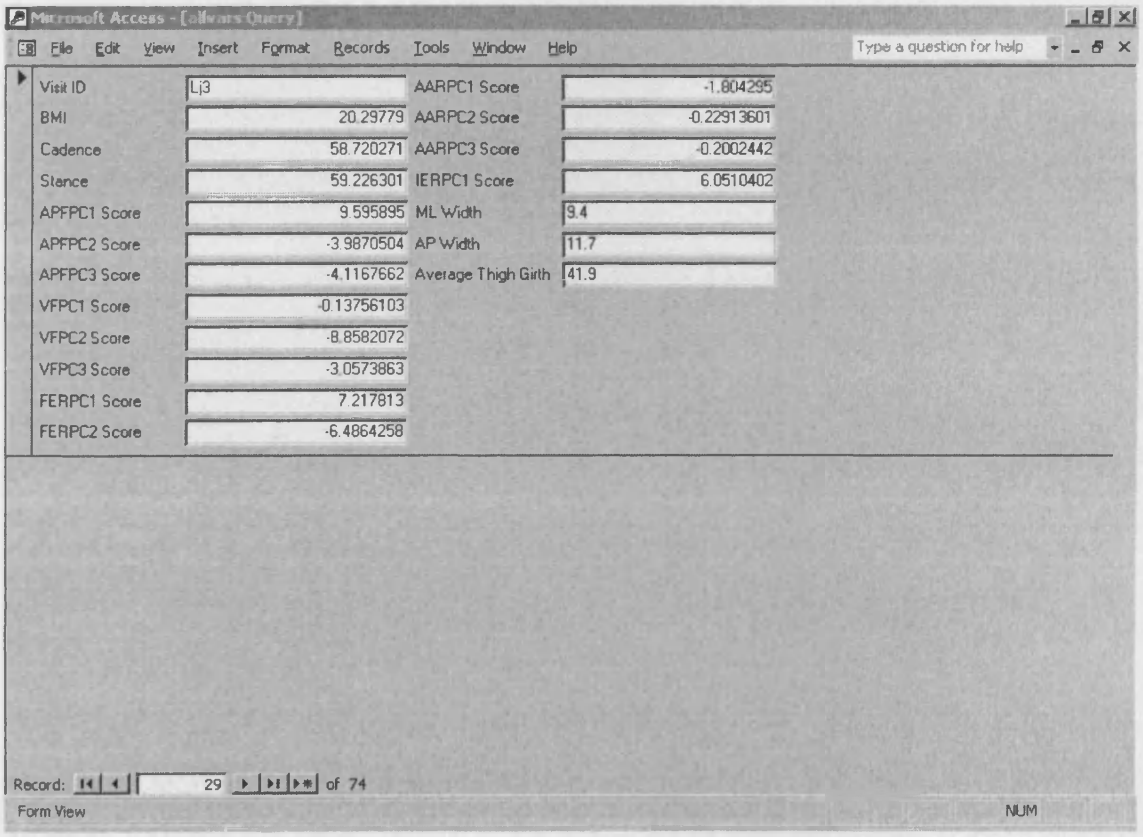
A database was created in Access 2000 (Microsoft) as shown in Figure 2.34 to store the patient details recorded in Table 2.1 and the set of variables listed in Table 2.8 that were produced during the data processing stage.

Table 2.8 Variables produced during data processing.

Variable	Variable Description
v_1	BMI
v_2	Cadence
v_3	Stance
v_4	APFPC1 Score
v_5	APFPC2 Score
v_6	APFPC3 Score
v_7	VFPC1 Score
v_8	VFPC2 Score
v_9	VFPC3 Score
v_{10}	FERPC1 Score
v_{11}	FERPC2 Score
v_{12}	AARPC1 Score
v_{13}	AARPC2 Score
v_{14}	AARPC3 Score
v_{15}	IERPC1 Score
v_{16}	ML Width
v_{17}	AP Width
v_{18}	Thigh Girth



(a)



(b)

Figure 2.34 Access database: (a) patient details and (b) final dataset.

This chapter has provided an in-depth description of the methods used to measure and process the knee function data. Conclusions from and further work based on the motion analysis technique and the PCA work are given in Chapter 8.

CHAPTER 3

THE CLASSIFICATION METHOD

This chapter describes in detail the new classification method that was introduced in Chapter 1. The classification method is based around the Dempster-Shafer theory of evidence (DST) and builds on the work of Safranek *et al.* (1990) and Gerig *et al.* (2000) who developed and utilised parts of the classification method in the areas of vision and medical image analysis respectively. Founded on the work of Dempster (1968) and Shafer (1976), the DST is a method that enables decision-making in the presence of ignorance (Safranek *et al.*, 1990). The chapter begins with an introduction to the notion of ignorance and to the DST and its associated terminology before moving on to describe the classification method in detail.

3.1 THE MEANING OF IGNORANCE

A measure of ignorance is present in many decision-making processes (Shafer and Pearl, 1990; Lipschitz and Strauss, 1997), presenting itself in several different forms (Smets, 1991; Lipschitz and Strauss, 1997). Smets (1991) defines three sources of ignorance as incompleteness, imprecision and uncertainty whilst Lipschitz and Strauss (1997) cite three sources of uncertainty as inadequate understanding, incompleteness and undifferentiated alternatives. Smets asserts that different sources of ignorance can be present simultaneously which is particularly evident in the decision making process associated with gait classification. Definitions of ignorance that arise in the context of gait classification will now be explained.

Occasionally, the information collated during a session in the motion analysis laboratory is **incomplete**. For example, in order to measure the three-dimensional ground reaction forces a subject must plant their foot within the boundaries of the force plate whilst walking along a walkway i.e. strike the force plate "cleanly". If the force plate is not struck cleanly during a walk then a repeat reading must be taken. Suppose then that a subject with osteoarthritis experiences extreme pain in their knee when walking and as a result tires very quickly. In this instance the subject's

condition prevents repeat readings from being taken. Whilst all other information is collated during the session, the ground-reaction force data is missing, and as such the data is incomplete.

Some of the information gathered during a session in the motion analysis laboratory is **imprecise**. As an example, let us consider the kinematic measurements. A well-documented subject within the motion analysis community is the quantification of errors associated with skin movement artefacts and the reconstruction of marker coordinates (Cappello *et al.*, 1997; Cappozzo *et al.*, 1996). The skin movement artefacts can result in errors in the flexion-extension, abduction-adduction and internal-external rotation knee angle measurements of the magnitude of 8°, 4° and 12° respectively (Cappozzo *et al.*, 1996). As such the information collected is complete (because the value is not missing) but imprecise because the value of the variable, v, lies somewhere in the interval [v – error, v + error].

A third source of ignorance encountered in gait classification is **uncertainty due to inadequate understanding**. This form of uncertainty is helpfully defined by Lipschitz and Strauss (1997, pp.151): "Decision makers are sometimes unable to act not because they lack information but because they are overwhelmed by the abundance of conflicting meanings that it conveys". This source of ignorance is perhaps the most prevalent source in the decision-making associated with gait classification. During a session in the motion analysis laboratory a vast amount of data that relates to a subject's knee function is collected (see Chapters 1 and 2). The relevance of each variable in discriminating normal and osteoarthritic knee function is often unknown *a priori*. Of these variables, some may support, not support or offer no evidence to a specific subject's correct classification.

Different methods exist for coping with the various forms of ignorance in decision-making processes. A summary of these methods, cited by Smets (1991), is given in Figure 3.1. This chapter introduces a new classification method that exploits one of these methods, namely the DST for its ability to deal with uncertainty.

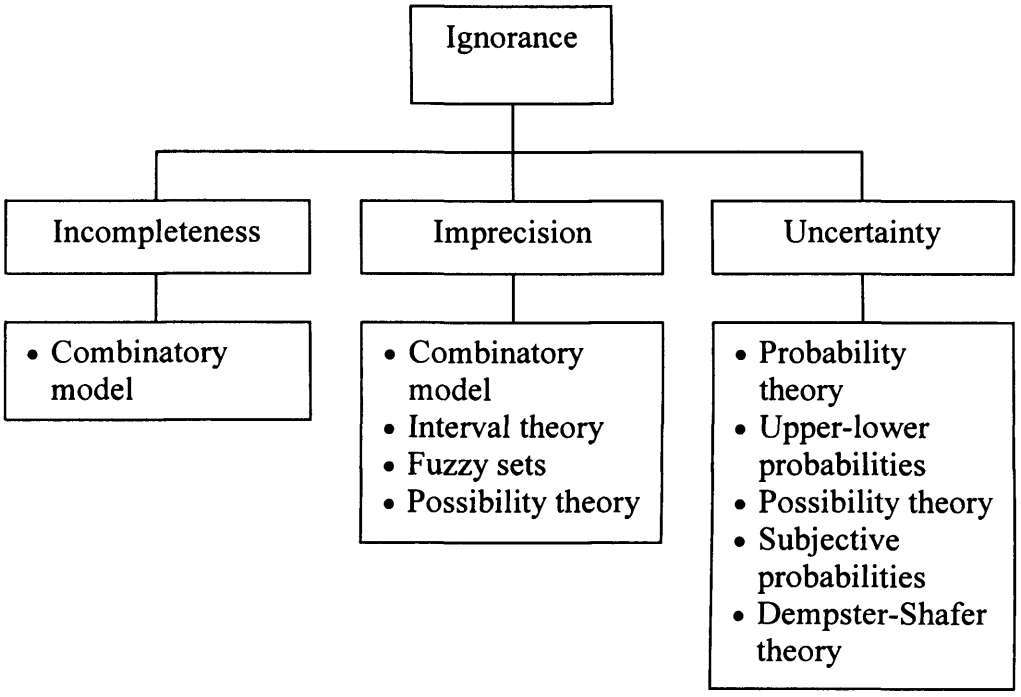


Figure 3.1 Methods that deal with the different forms of ignorance (Smets, 1991).

3.2 THE DEMPSTER-SHAFER THEORY (DST)

Decision-making processes under uncertainty are built upon the foundations of probability theory, and in general upon the Bayesian decision theory (Ng and Abramson, 1990; Beynon *et al.*, 2000). Indeed, Shafer (Shafer and Pearl, 1990, pp.61) comments that "...Bayesian decision theory has come to define what is meant by a general theory of decision making under uncertainty....its familiarity makes it a reference point to which other theories must relate. Any general account of decision making must define itself relative to Bayesian decision theory. Any such account must explain why it does not do things that Bayesian decision theory does or why it does some things differently."

Briefly, Bayesian decision theory is based around Bayes' theorem:

$$p(w_j|v_i) = \frac{p(v_i|w_j)p(w_j)}{p(v_i)} \qquad (3.1)$$

where $p(w_j|v_i)$ is the posterior probability, which is the probability of the outcome being w_j given that v_i has been measured; $p(v_i|w_j)$ is the conditional probability of obtaining the measurement v_i given that the outcome is w_j; $p(w_j)$ is the prior probability of the outcome being w_j and $p(v_i) = \sum_{j=1}^{n} p(v_i|w_j)p(w_j)$ (Duda, Hart and Stork, 2001). In most practical situations the prior and conditional probabilities are not known and must be estimated from training data. In cases where the sample size is small this can lead to poor estimates of the prior and conditional probabilities (Denoeux, 1997).

A further problem with the Bayesian approach occurs when there is a lack of knowledge relating to the exact number of outcomes (Denoeux, 1997; Beynon *et al.*, 2000). Let us consider an example which explores the hypothesis "Lianne lives in Donald Street, Cardiff" = w_1 (adapted from Beynon *et al.*, 2000). The finite set of possible outcomes is dependent upon the number of streets that are known to exist in Cardiff. If only 10 streets are identifiable then the finite set of possible outcomes is $\{w_1, w_2,...,w_{10}\}$ where w_j ($j \neq 1$) represent other roads in Cardiff. Using a Bayesian approach the probabilities are evenly distributed over this set, which implies that $p(w_1) = 0.1$. However, if 1000 streets are identifiable then the finite set of possible outcomes is $\{w_1, w_2,...,w_{1000}\}$ and $p(w_1) = 0.001$. The example shows that the prior probability is also highly dependent on the finite set of possible outcomes (Beynon *et al.*, 2000).

The DST is founded on the work of Dempster (1968) and Shafer (1976). Shafer (Shafer and Pearl, 1990, pp.473) comments that DST "provides a non-Bayesian way of using mathematical probability to quantify subjective judgements. Whereas a Bayesian assesses probabilities directly for the answer to a question of interest, a belief function user assesses probabilities for related questions and then considers the implications of these probabilities for the question of interest."

The DST comprises two main elements: the assignment of belief values to different hypotheses and the combination of belief values (Shafer and Pearl, 1990). These two elements are now expounded in sections 3.2.1 and 3.2.2 respectively. In both sections

the differences between the DST and the Bayesian approach are discussed. This introduction to the basic concepts in DST is based on Safranek *et al.* (1990)

3.2.1 Assignment of belief values

A frame of discernment (FOD) is defined as a finite set of mutually exclusive and exhaustive elementary hypotheses or elementary propositions $\{\phi_i\}$, and is denoted by Θ. The number of elementary hypotheses in Θ is symbolized by $|\Theta|$. There are $2^{|\Theta|}$ possible subsets of Θ and these are defined as hypotheses or propositions, $\{\varphi_i\}$, ($i = 1,\ldots 2^{|\Theta|}$). The set of hypotheses is denoted by the power set 2^{Θ} and contains the FOD Θ, and the null hypothesis or empty set, $\emptyset$. A binary frame of discernment (BFOD) is a FOD that contains only two elementary hypotheses, $\{\phi_1\}$ and $\{\phi_2\}$, i.e. $\Theta = \{\phi_1, \phi_2\}$. There are $2^{|\Theta|} = 2^2 = 4$ subsets of Θ - the hypotheses $\{\phi_1\}$, $\{\phi_2\}$, $\{\phi_1, \phi_2\} = \Theta$ and $\emptyset$ respectively.

In DST, each hypothesis (subset of Θ) is assigned a degree of belief based upon the available evidence. This degree of belief is called the basic probability assignment (*bpa*) or probability mass function, $m(.)$. The *bpa* "expresses the degree to which the evidence confirms a hypothesis" (Ji and Marefat, 2003, pp.1362). A *bpa* is defined as follows (Safranek *et al.*, 1990):

$$0 \le m(.) \le 1$$

$$m(\emptyset) = 0 \qquad (3.2)$$

$$\sum_{i=1}^{n} m(\varphi_i) = 1$$

Each hypothesis $\{\varphi_i\}$, ($i = 1,\ldots, n$) that has $m(\varphi_i) > 0$ is defined as a focal element. The focal elements and their associated *bpa* define a body of evidence (BOE).

To further understanding of the DST and to identify differences between DST and the Bayesian approach an example, which considers the question "Who killed top secret agent Parkes in Cardiff last week?", is expounded (adapted from Beynon *et al.*,

2000)). It is known that the assassin is one of the three villainous gangsters Jones, Cruise and Bennett. Thus, the FOD contains the elementary hypotheses "the assassin is Jones", {Jones}; "the assassin is Cruise", {Cruise}; and "the assassin is Bennett", {Bennett}; i.e. Θ = {Jones, Cruise, Bennett}. The number of elementary hypotheses, $|\Theta| = 3$ and the number of hypotheses = $2^3 = 8$. The hypotheses are recorded in Table 3.1.

Table 3.1 All possible hypotheses for assassin example.

Hypothesis
{Jones}
{Cruise}
{Bennett}
{Jones, Cruise}
{Jones, Bennett}
{Cruise, Bennett}
{Jones, Cruise, Bennett} = Θ
Ø

Witnesses at the crime scene believe with 80% certainty that the assassin was male. This gives a *bpa*, say m_1 of 0.8 to the focal element {Cruise, Bennett} i.e. m_1({Cruise, Bennett}) = 0.8. Using DST no presumptions are made about the remaining probability mass and so it is assigned to the entire FOD, Θ = {Jones, Cruise, Bennett} i.e. m_1({Jones, Cruise, Bennett}) = $m(\Theta)$ = 0.2.

It is in this assignment of beliefs that DST allows for uncertainty. In DST, probability masses are not only assigned to elementary hypotheses, but to each hypothesis or subset of Θ (Safranek *et al.*, 1990). The probability mass is free to move to any element of the subset. Consequently the probability mass assigned to Θ, $m(\Theta)$, represents uncertainty, because the mass is free to move to any elementary hypothesis in the FOD (Safranek *et al.* 1990).

It is also at this point that the greatest conflict between Bayesian probability and DST lies. In contrast to DST, in Bayesian probability, probability masses are only assigned to elementary hypotheses and not to subsets of Θ, implying that Bayesian probability cannot represent uncertainty (Safranek *et al.*, 1990). In the Bayesian approach, the probability that is not committed to a hypothesis, say a, must be given to its negation, $\neg a$ i.e. $p(a) + p(\neg a) = 1$ (Ng and Abramson, 1990). For example, using the above assassin example, if $p(\text{man}) = 0.8$, then it is required that $p(\text{woman}) = 0.2$. In contrast, using DST, since belief values can be assigned to any subset of Θ, the belief values assigned to an elementary hypotheses, $\{a\}$ does not require that the remaining belief value be assigned to its negation $\{\neg a\}$ (Beynon *et al.*, 2000; Ng and Abramson, 1990). For example, using the assassin example, in the absence of any other information, $m_1(\{\text{man}\}) = 0.8$ does not imply that $m_1(\{\text{female}\}) = 0.2$, but that $m_1(\{\text{male, female}\}) = 0.2$. If $m(\Theta) = 0$ the resulting *bpa* is Bayesian since in this instance probability masses are only assigned to elementary hypotheses. As such, DST is said to be a generalisation of Bayesian probability theory (Shafer and Pearl, 1990).

3.2.2 Combination of belief values

The second element of DST is the combination of individual independent BOE. This is achieved using Dempster's rule of combination. The combination of two independent BOE m_1 and m_2 is defined $(m_1 \oplus m_2)$. If $m_c = (m_1 \oplus m_2)$ then

$$m_c(\varphi) = (m_i \oplus m_j)(\varphi) = \kappa \sum_{s_i \cap s_j = \varphi} m_i(s_i) m_j(s_j)$$

$$\kappa^{-1} = 1 - \sum_{s_i \cap s_j = \emptyset} m_i(s_i) m_j(s_j) \qquad (3.3)$$

where κ is a normalisation factor, s_i and s_j are focal elements from the BOE $m_i(.)$ and $m_j(.)$, and $m_c(\varphi)$ is a bpa if $\kappa^{-1} \neq 0$ (Safranek *et al.*, 1990). When $\kappa^{-1} = 0$, $m_c(\varphi)$ does not exist and m_i and m_j are wholly conflicting pieces of evidence. When $\kappa^{-1} = 1$, m_i and m_j are fully concordant. (Mellouli and Elouedi, 1997) As such, κ can be understood as a measure of conflict that exists between the sources of evidence (Beynon *et al.* 2000). Dempster's rule is commutative and associative and hence

$m_c(\varphi)$ does not depend on the order in which the evidence is combined (Gerig *et al.* 2000).

This combination of beliefs is explained further using the above assassin example (adapted from Beynon *et al.*, 2000). Two further pieces of evidence have been uncovered relating to the murder of agent Parkes:

1. Samples of dark brown hair were found at the crime scene
2. The murder weapon was found in Donald Street

The *bpa* relating to these two pieces of evidence are given in Table 3.2 as m_2 and m_3 respectively (values taken from Beynon *et al.*, 2000). Using equation (3.3) and the *bpa* recorded in Table 3.2, an example of combining the evidence $m_2 \oplus m_3 = m_4$, is given:

$$m_4(\{Jones, Cruise\}) = (m_2 \oplus m_3)(\{Jones, Cruise\}) = \frac{\sum_{s_2 \cap s_3 = \{\text{Jones,Cruise}\}} m_2(s_2) m_3(s_3)}{1 - \sum_{s_2 \cap s_3 = \emptyset} m_2(s_2) m_3(s_3)} \tag{3.4}$$

Calculating the numerator from equation (3.4):

$$\begin{aligned}\sum_{s_2 \cap s_3 = \{\text{Jones,Cruise}\}} m_2(s_2) m_3(s_3) &= [m_2(\{\text{Jones, Cruise}\}) \times m_3(\{\text{Jones, Cruise}\})] + \ldots \\ &+ [m_2(\{\text{Jones, Cruise}\}) \times m_3(\{\text{Jones, Cruise, Bennett}\})] + \ldots \\ &+ [m_3(\{\text{Jones, Cruise}\}) \times m_2(\{\text{Jones, Cruise, Bennett}\})]\end{aligned}$$

$$= [0.1 \times 0.1] + [0.1 \times 0.1] + [0.1 \times 0.2] = 0.04 \tag{3.5}$$

Calculating part of the denominator from equation (3.4):

$$\sum_{s_2 \cap s_3 = \emptyset} m_2(s_2) m_3(s_3) = [m_2(\{\text{Jones}\}) \times m_3(\{\text{Cruise}\})] + \ldots$$
$$+ [m_2(\{\text{Jones}\}) \times m_3(\{\text{Bennett}\})] + \ldots$$
$$+ [m_3(\{\text{Jones}\}) \times m_2(\{\text{Cruise}\})] + \ldots$$
$$+ [m_3(\{\text{Jones}\}) \times m_2(\{\text{Bennett}\})] + \ldots$$
$$+ [m_2(\{\text{Cruise}\}) \times m_3(\{\text{Bennett}\})] + \ldots$$
$$+ [m_3(\{\text{Cruise}\}) \times m_2(\{\text{Bennett}\})] + \ldots$$
$$+ [m_2(\{\text{Jones, Cruise}\}) \times m_3(\{\text{Bennett}\})] + \ldots$$
$$+ [m_3(\{\text{Jones, Cruise}\}) \times m_2(\{\text{Bennett}\})] + \ldots$$
$$+ [m_2(\{\text{Jones, Bennett}\}) \times m_3(\{\text{Cruise}\})] + \ldots$$
$$+ [m_3(\{\text{Jones, Bennett}\}) \times m_2(\{\text{Cruise}\})] + \ldots$$
$$+ [m_2(\{\text{Cruise, Bennett}\}) \times m_3(\{\text{Jones}\})] + \ldots$$
$$+ [m_3(\{\text{Cruise, Bennett}\}) \times m_2(\{\text{Jones}\})]$$

$$= [0.1 \times 0.1] + [0.1 \times 0.1] + [0.2 \times 0.1] + [0.2 \times 0.2] + [0.1 \times 0.1] + [0.1 \times 0.2] + \ldots$$
$$+ [0.1 \times 0.1] + [0.1 \times 0.2] + [0.2 \times 0.1] + [0.3 \times 0.1] + [0.1 \times 0.2] + [0.1 \times 0.1]$$
$$= 0.22 \qquad (3.6)$$

Substituting equations (3.5) and (3.6) into (3.4) gives

$$m(\{\text{Jones, Cruise}\}) = \frac{0.04}{1 - 0.22} = 0.051 \qquad (3.7)$$

Using equation (3.3) the allocation of the combined evidence $m_4 = m_2 \oplus m_3$ for the remainder of the hypotheses is given in Table 3.2.

Table 3.2 Combination of belief values for assassin example

	∅	{Jones}	{Cruise}	{Bennett}	{Jones, Cruise}	{Jones, Bennett}	{Cruise, Bennett}	{Jones, Cruise, Bennett}
m_2	0	0.1	0.1	0.2	0.1	0.2	0.1	0.2
m_3	0	0.2	0.1	0.1	0.1	0.3	0.1	0.1
m_4	0	0.282	0.128	0.282	0.051	0.180	0.32	0.026

3.3 PREVIOUS WORK

Safranek *et al.* (1990) and Gerig *et al.* (2000) utilised aspects of DST in studies in the areas of vision and medical image analysis respectively and their works highlight the potential of DST as a classification tool. Their contributions are now discussed in detail.

3.3.1 Safranek *et al.* (1990)

Safranek *et al.* (1990) developed a method using DST for object detection and verification tasks in robot vision. Their task of verifying the position of objects based on camera images of these objects can be summarised as following:

i. Collect measurements of the object from the camera image
ii. Determine the extent to which each measurement supports the existence of the object
iii. Combine the evidence from all the measurements to make a final decision as to whether or not the object is at its predicted position

Because of the uncertain nature of their measurements, Safranek *et al.* (1990) used DST to assign levels of support to each measurement and subsequently to combine these individual pieces of evidence (steps ii and iii above). They used BFOD and defined the elementary hypotheses as "the object is at its predicted position", $\{x\}$, and "the object is not at its predicted position", $\{\neg x\}$, i.e. $\Theta = \{x, \neg x\}$.

The first stage of their method is to standardise each input variable (v). Each input variable is mapped to a value on a scale of 0 to 1. The transformed input variable is defined the confidence factor, $cf(v)$. Safranek *et al.* (1990) state that $cf(v)$, must satisfy the following criteria:

i. $cf(v)$ is an increasing function
ii. $cf(v) = 1$ if the measurement implies certainty in $\{x\}$
iii. $cf(v) = 0$ if the measurement implies certainty in $\{\neg x\}$
iv. $cf(v) = 0.5$ if the measurement favours neither $\{x\}$ nor $\{\neg x\}$

The confidence factor representing each input variable is subsequently transformed into a BOE. Given that the subsets of Θ are the hypotheses $\{x\}$, $\{\neg x\}$, $\{x, \neg x\}$ and $\emptyset$, then the associated *bpa* are:

i.	$m(\{x\})$	– the degree of belief in $\{x\}$
ii.	$m(\{\neg x\})$	– the degree of belief in $\{\neg x\}$
iii.	$m(\{x, \neg x\}) = m(\Theta)$	– the degree of belief in $\{x, \neg x\}$
iv.	$m(\emptyset) = 0$	– the degree of belief in neither x nor $\neg x$

Safranek *et al.* (1990) define these *bpa* as:

$$m(\{x\}) = \frac{B}{1-A} cf(v) - \frac{AB}{1-A} \tag{3.4}$$

$$m(\{\neg x\}) = \frac{-B}{1-A} cf(v) + B \tag{3.5}$$

and

$$m(\Theta) = 1 - m(\{x\}) - m(\{\neg x\}) = \frac{1-A-B}{1-A} \tag{3.6}$$

where A depicts the dependence of $m(\{x\})$ on the confidence factor and B the maximal support assigned to $m(\{x\})$ or $m(\{\neg x\})$. Subsequently, the *bpa* corresponding to each individual measurement are combined using Dempster's rule (see equation 3.3) to produce a combined BOE (BOE_c) defined by the *bpa* $m_c(\{x\})$, $m_c(\{\neg x\})$ and $m_c(\Theta)$.

Finally, a decision is made based on the BOE_c:

i. If $m_c(\{x\}) > m_c(\{\neg x\})$ then the outcome is considered to be x
ii. If $m_c(\{\neg x\}) > m_c(\{x\})$ then the outcome is considered to be $\neg x$

3.3.2 Gerig *et al.* (2000)

Due to the uncertainty associated with the measurements used to discriminate between lesions and static tissue, Gerig *et al.* (2000) applied DST to the detection of brain lesions in subjects with multiple sclerosis. Following Safranek *et al.* (1990), Gerig *et al.* (2000) limited their study to a BFOD, which they defined as Θ = {lesions, ¬ lesions}.

3.4 NEW CLASSIFICATION METHOD FOR OSTEOARTHRITIC (OA) AND NORMAL (NL) KNEE FUNCTION

Building on the work of Safranek *et al.* (1990) and Gerig *et al.* (2000), this section introduces a new classification method for OA and NL knee function. Following the work of Safranek *et al.* (1990), the classification method is restricted to BFOD. The BFOD, Θ, contains two antagonistic elementary hypotheses "a subject has OA knee function" denoted {OA} and a "subject has NL knee function" denoted {NL}, i.e. Θ = {OA, NL}. There are $2^{|\Theta|} = 2^2 = 4$ subsets of Θ - the hypotheses {OA}, {NL}, {OA, NL} = Θ and $\emptyset$.

The new classification method comprises a number of stages:

i. Conversion of input variables into confidence factors
ii. Conversion of confidence factors to BOE
iii. Combination of individual BOE
iv. Visualisation of BOE using simplex plots
v. Classification based on the final combined BOE, BOE_c

These individual stages are now discussed in the following subsections, throughout which the reader is directed to Figure 3.2, which illustrates the interconnection of the individual stages.

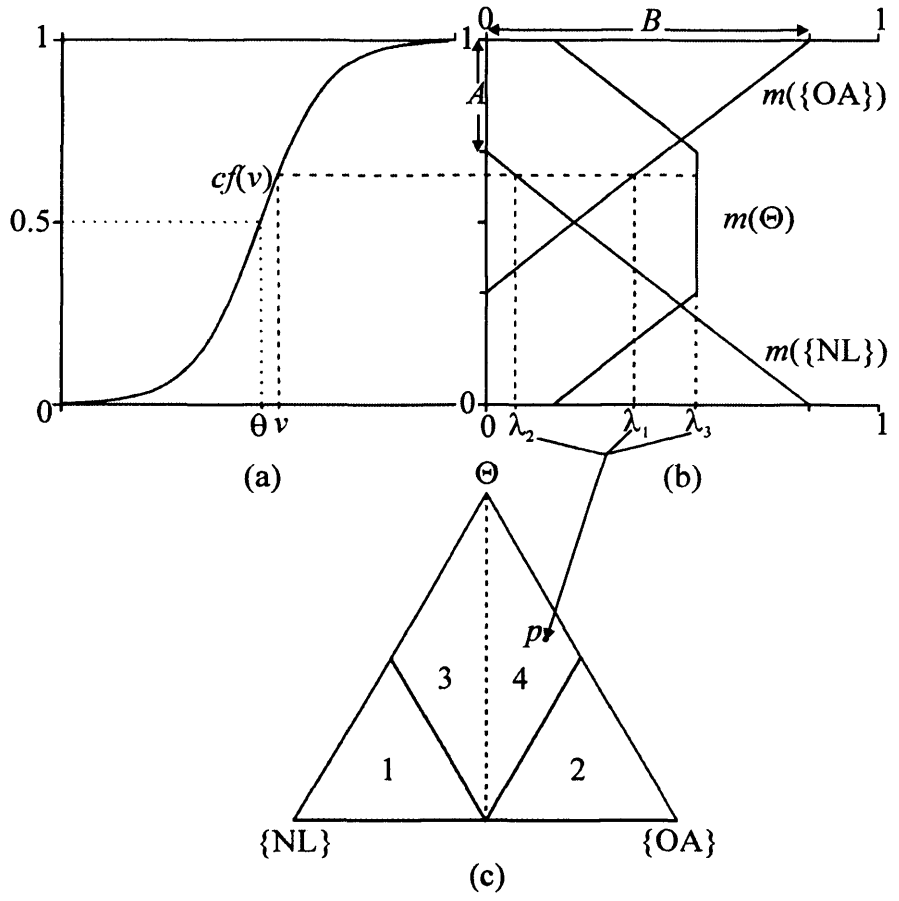


Figure 3.2 The classification method showing the interaction of its three main stages. (a) Conversion of input variable, v, into confidence factor $cf(v)$ using the sigmoid function. θ is the value of v for which $cf(v) = 0.5$. (b) Conversion of confidence factor into body of evidence (BOE) comprising the bpa $m(\{OA\}) = \lambda_1$, $m(\{NL\}) = \lambda_2$ and $m(\Theta) = \lambda_3$. A and B are the DS control parameters. (c) Conversion of the BOE into its simplex coordinate, denoted by the point p (adapted from Beynon *et al.*, 2002). The simplex plot is divided into four regions: 1 denotes the dominant NL classification region; 2 denotes the dominant OA classification region; 3 denotes the non-dominant NL classification region and 4 denotes the non-dominant OA classification regions. The dotted vertical line is the decision boundary.

3.4.1 Conversion of input variables into confidence factors

The first stage of the classification method is to standardise each input variable, v. This standardisation process ensures that the input variables are in an appropriate form for the *bpa* (Gerig *et al.* 2000). Following the work of Safranek *et al.* (1990), each input variable is mapped to a value on a scale of 0 to 1. The transformed input variable is defined the confidence factor, $cf(v)$, and represents a level of confidence in (or not in) the variable's support to a subject's knee function being OA. The greater the value of the confidence factor the greater the support for a subject's knee function being OA and vice versa. The confidence factor $cf(v)$, must satisfy the following criteria (adapted from Safranek *et al.*, 1990):

i. $cf(v)$ is a monotonic function
ii. $cf(v) = 1$ if the measurement implies certainty in {OA}
iii. $cf(v) = 0$ if the measurement implies certainty in {NL}
iv. $cf(v) = 0.5$ if the measurement favours neither {OA} nor {NL}

Safranek *et al.* (1990) comment that the choice of such a function is application specific. Gerig *et al.* (2000) suggest that one possible function to transform the input variables into confidence factors is the sigmoid function

$$cf(v) = \frac{1}{1 + e^{-k(v-\theta)}} \tag{3.7}$$

where k describes the gradient of the confidence function, $cf(v)$, and θ is the value of v which produces a confidence value, $cf(v) = 0.5$ as shown in Figure 3.2a. The sigmoid function is commonly used as an activation function within Artificial Neural Networks including applications to gait analysis studies (e.g. Holzreiter and Köhle, 1993).

A sigmoid confidence function with a positive gradient (positive association) implies that a large v measurement offers more support to {OA} ($cf(v) \rightarrow 1$), whilst a small v measurement offers more support to {NL} ($cf(v) \rightarrow 0$) (Figure 3.3a). Conversely, a negative gradient (negative association) implies that a large v measurement offers

more support to {NL} whilst a small v measurement offers more support to {OA} (Figure 3.3b).The nature of association of each k can be deduced from knowledge of its associated input variable.

The absolute value of k indicates the range of the measurement of the input variable that lies around the middle confidence value $cf(v) = 0.5$. Smaller absolute values of k increase the range of v for which $cf(v)$ is near 0.5, thus restricting the majority of v measurements to be assigned a $cf(v)$ value of near 0.5 (Figure 3.3c). Conversely, a greater absolute value of k, decreases the range of v for which $cf(v)$ is near 0.5, which implies that the majority of v measurements transform towards the extreme $cf(v)$ values of 0 or 1 (Figure 3.3d)

The θ variable is the value of v for which the evidence supporting {OA} is equal to the evidence supporting {NL} when $cf(v) = 0.5$ as shown in Figure 3.4. Movement to either side of this value leads to an increase in the level of support to either {OA} or {NL}.

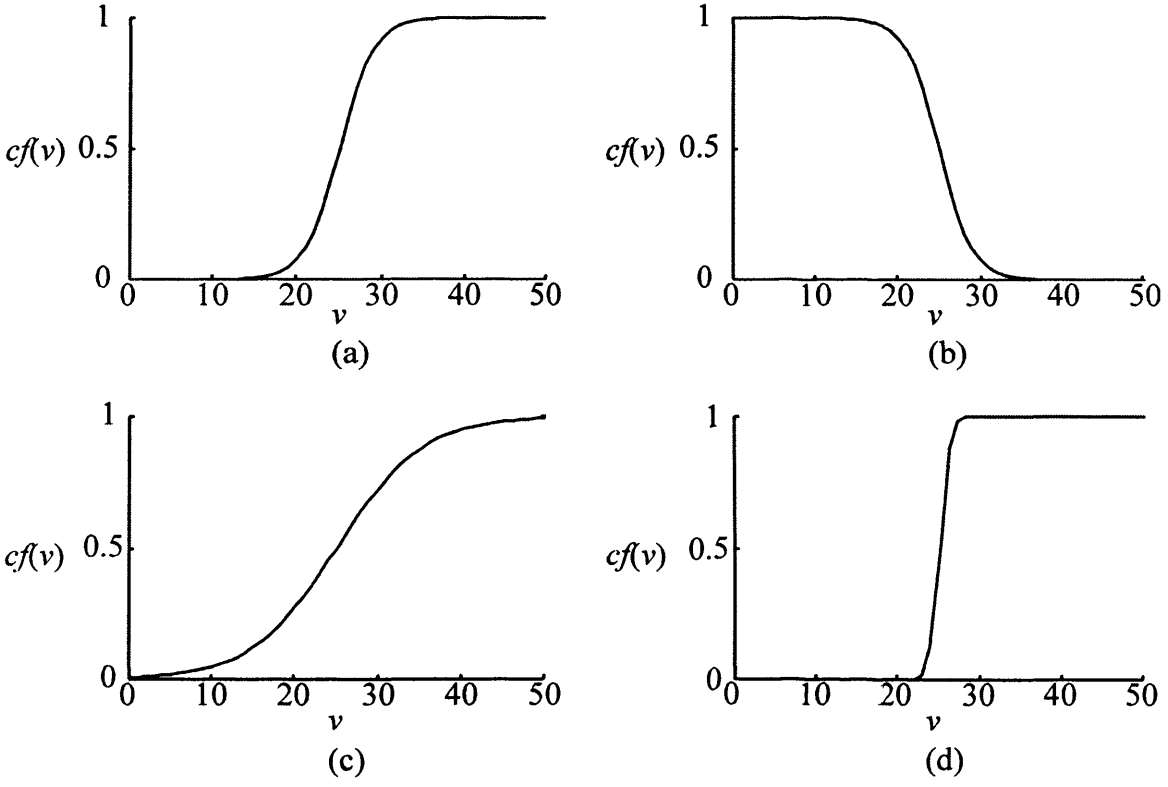


Figure 3.3 Influence of k on confidence factor (a) positive association ($k = 0.25$) (b) negative association ($k = -0.25$) (c) small absolute value of k ($k = 0.2$) (d) large absolute value of k ($k = 2$)

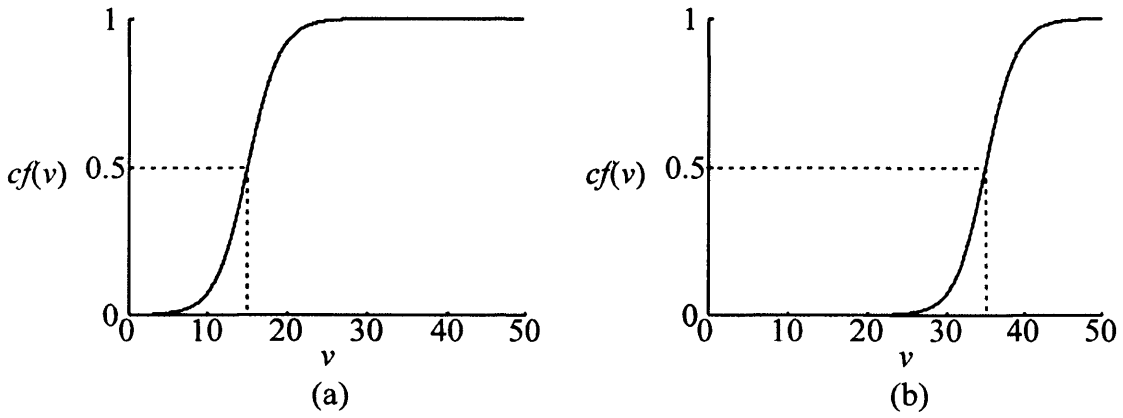


Figure 3.4 Dependence of confidence factor on θ (a) $\theta = 15$ (b) $\theta = 35$

3.4.2 Conversion of confidence factors to BOE using DST

The confidence factor representing each input variable is subsequently transformed into a BOE; a set of belief measures established within the context of DST (see Figure 3.2b). Given that the subsets of Θ are the hypotheses {OA}, {NL}, {OA, NL} = Θ and $\emptyset$, then the associated *bpa* are:

i. $m(\{OA\})$ – the degree of belief in {OA}
ii. $m(\{NL\})$ – the degree of belief in {NL}
iii. $m(\{OA, NL\}) = m(\Theta)$ – the degree of belief in {OA, NL}
iv. $m(\emptyset) = 0$ – the degree of belief in neither {OA} nor {NL}

The value $m(\{OA, NL\}) = m(\Theta)$ is the associated ignorance because it represents the value that cannot be given to {OA} or {NL} explicitly. From Safranek *et al.* (1990) these *bpa* are defined as:

$$m(\{OA\}) = \frac{B}{1-A} cf(v) - \frac{AB}{1-A} \tag{3.8}$$

$$m(\{NL\}) = \frac{-B}{1-A} cf(v) + B \tag{3.9}$$

and

$$m(\Theta) = 1 - m(\{\text{OA}\}) - m(\{\text{NL}\}) = \frac{1 - A - B}{1 - A} \tag{3.10}$$

where A depicts the dependence of $m(\{\text{OA}\})$ on the confidence factor and B the maximal support assigned to $m(\{\text{OA}\})$ or $m(\{\text{NL}\})$. Imperatively, if either $m(\{\text{OA}\})$ or $m(\{\text{NL}\})$ are less than zero their values are set to zero. The $m(\Theta)$ is then calculated subsequent to making these changes.

3.4.3 Combination of individual BOE

If there is more than one input variable characterising a subject, a respective number of number of BOE will be constructed. Each BOE offers positive or negative evidence to support the classification of a subject to a hypothesis. The Dempster's rule of combination, which assumes that the input variables are independent, is used to combine the individual BOE into a final combined BOE (BOE_c). In the case of BFOD, where Θ = {OA, NL} this rule is given by the three following formulaic expressions (based on two independent BOE $m_i(.)$ and $m_j(.)$) (Gerig *et al.*, 2000):

$$(m_i \oplus m_j)(\{\text{OA}\}) = m_c(\{\text{OA}\}) = \frac{m_i(\{\text{OA}\})m_j(\{\text{OA}\}) + m_j(\{\text{OA}\})m_i(\Theta) + m_i(\{\text{OA}\})m_j(\Theta)}{1 - (m_i(\{\text{NL}\})m_j(\{\text{OA}\}) + m_i(\{\text{OA}\})m_j(\{\text{NL}\}))} \tag{3.11}$$

$$(m_i \oplus m_j)(\{\text{NL}\}) = m_c(\{\text{NL}\}) = \frac{m_i(\{\text{NL}\})m_j(\{\text{NL}\}) + m_j(\Theta)m_i(\{\text{NL}\}) + m_j(\{\text{NL}\})m_i(\Theta)}{1 - (m_i(\{\text{NL}\})m_j(\{\text{OA}\}) + m_i(\{\text{OA}\})m_j(\{\text{NL}\}))} \tag{3.12}$$

$$(m_i \oplus m_j)(\Theta) = m_c(\Theta) = 1 - m_c(\{\text{OA}\}) - m_c(\{\text{NL}\}) \tag{3.13}$$

The combination of individual BOE to form the BOE_c is illustrated in Figure 3.5. The BOE_c comprises the same three focal elements as present in the individual BOE namely {OA}, {NL} and Θ.

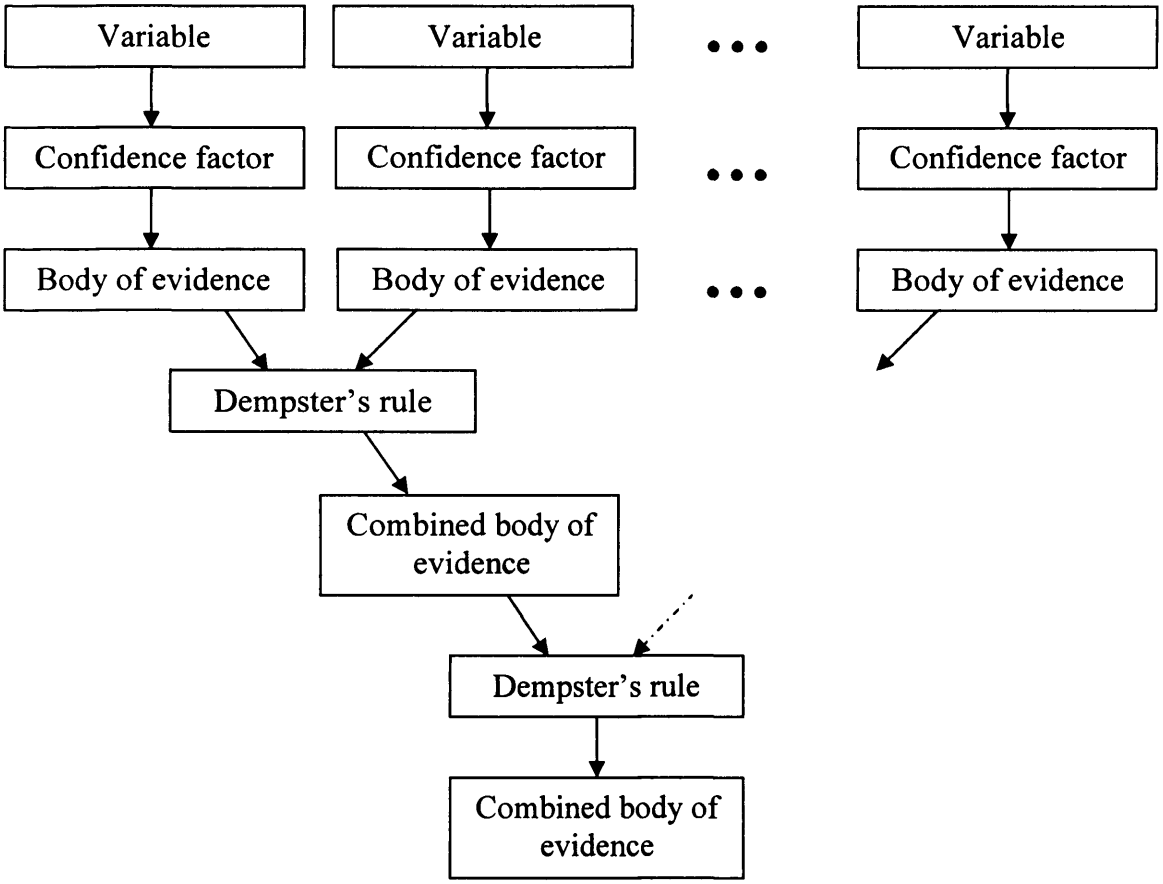


Figure 3.5 Evidence accumulation. Combination of individual bodies of evidence using Dempster's rule (adapted from Gerig *et al.*, 2000, pp.36).

3.4.4 Visualisation of BOE using simplex plots

Following the work of Beynon *et al.* (2002), a simplex coordinate is used to represent the BOE as a single point in a simplex plot (see Figure 3.2c). A mathematical derivation is now given to explain this conversion of the BOE into its simplex coordinate (see Silvester and Ferrari, 1996; Coxeter, 1969).

Figure 3.6 shows an equilateral triangle $\triangle A_1A_2A_3$ with area denoted by S. A point p lies somewhere within this triangle and divides it into three sub-triangles $\triangle A_1A_3p$, $\triangle A_2A_3p$, and $\triangle A_1A_2p$ which have a common vertex p and areas S_k, k = 1, 2, 3, respectively as shown in Figure 3.7. It follows that

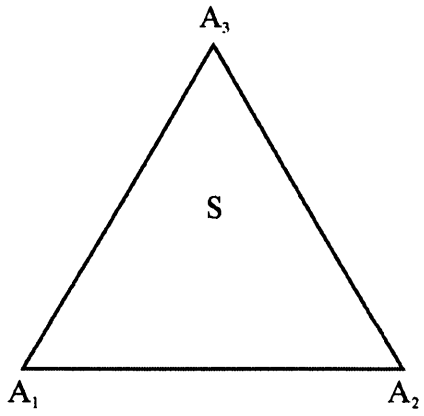


Figure 3.6 Equilateral triangle $\triangle A_1A_2A_3$ with area denoted by S.

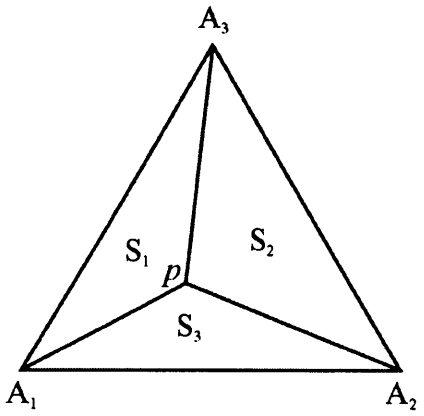


Figure 3.7 Equilateral triangle $\triangle A_1A_2A_3$ divided into three sub-triangles $\triangle A_1A_3p$, $\triangle A_2A_3p$, and $\triangle A_1A_2p$ which have a common vertex p and areas S_k, $k = 1, 2, 3$, respectively.

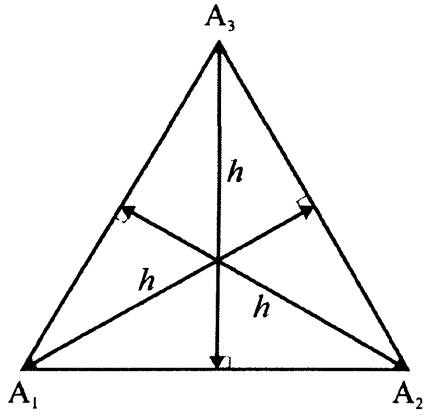


Figure 3.8 h is the shortest distance from each vertex to its respective opposite edge.

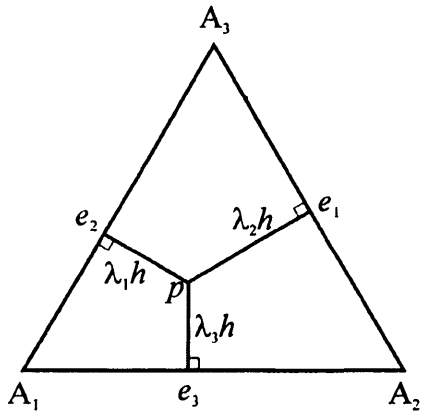


Figure 3.9 The shortest distance from p to the three edges.

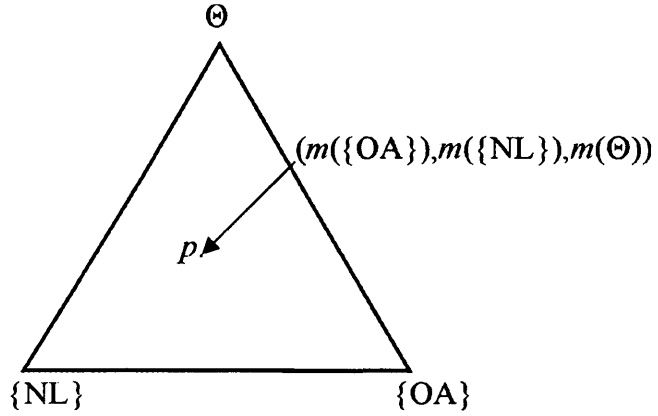


Figure 3.10 The simplex coordinate of p is (m({OA}), m({NL}), m(Θ)).

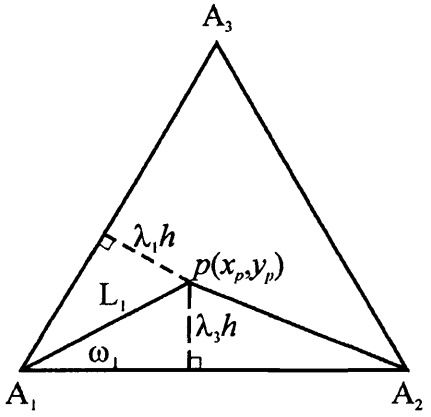


Figure 3.11 Angle ω_1 is defined as the angle between the lines A_1A_2 and A_1p. L_1 is the length of the line A_1p.

$$S = \sum_{k=1}^{3} S_k \tag{3.14}$$

Dividing through by S gives

$$\sum_{k=1}^{3} \frac{S_k}{S} = 1 \tag{3.15}$$

The shortest distance from each vertex to its respective opposite edge is given by h as shown in Figure 3.8. The shortest distance from p to the edge A_1A_2, is the length of the perpendicular bisector of A_1A_2 passing through p, the line pe_3. The length of pe_3 is a fraction, λ_3 ($0 \leq \lambda_3 \leq 1$), of the length h, as shown in Figure 3.9.

Since triangles $\triangle A_1A_2A_3$ and $\triangle A_1A_2p$ have a common base (Figure 3.7), it follows that the ratio of the area of triangle $\triangle A_1A_2A_3$ to the area of triangle $\triangle A_1A_2p$ is equal to the ratio of the height of triangle $\triangle A_1A_2A_3$ to the height of triangle $\triangle A_1A_2p$, i.e.,

$$\frac{\lambda_3 h}{h} = \frac{S_3}{S} \tag{3.16a}$$

Similarly,

$$\frac{\lambda_1 h}{h} = \frac{S_1}{S} \tag{3.16b}$$

and

$$\frac{\lambda_2 h}{h} = \frac{S_2}{S} \tag{3.16c}$$

Substituting these three results into equation (3.15) gives

$$\lambda_1 + \lambda_2 + \lambda_3 = 1 \tag{3.17}$$

The three numbers, λ_1, λ_2 and λ_3, specify the point p uniquely within the equilateral triangle $\triangle A_1A_2A_3$, and $(\lambda_1, \lambda_2, \lambda_3)$ are known as the simplex coordinate of p.

Comparing equation (3.10) with equation (3.17) indicates that the three belief measures can be represented using a simplex coordinate if we define

$$m(\{OA\}) = \lambda_1 \tag{3.18a}$$

$$m(\{NL\}) = \lambda_2 \tag{3.18b}$$

$$m(\Theta) = \lambda_3 \tag{3.18c}$$

Therefore, in the simplex plot a point p exists within an equilateral triangle such that the least distance from p to each side of the equilateral triangle are in the same proportion to the values $m(\{OA\})$, $m(\{NL\})$ and $m(\Theta)$. Briefly, the nearer a simplex coordinate is to a specific vertex the more association the BOE has to that subset of the frame of discernment. Thus, within the simplex plot the three belief measures can be represented using the simplex coordinate $(m(\{OA\}), m(\{NL\}), m(\Theta))$ provided that the vertices A_1, A_2 and A_3 correspond to $\{NL\}$,$\{OA\}$ and Θ respectively, as shown in Figure 3.10.

The vertices A_1, A_2 and A_3 have the simplex coordinates (0,1,0), (1,0,0) and (0,0,1) respectively. In the limits therefore, if the point p lies at the bottom right vertex of the triangle, $m(\{OA\}) = 1$ and $m(\{NL\}) = m(\Theta) = 0$; if the point p lies at the bottom left vertex of the triangle, $m(\{NL\}) = 1$ and $m(\{OA\}) = m(\Theta) = 0$; if the point p lies at the top vertex of the triangle, $m(\Theta) = 1$ and $m(\{OA\}) = m(\{NL\}) = 0$.

Let the angle between the edge A_1A_2 and the line A_1p be ω_1 and let the length of the line A_1p be L_1, as shown in Figure 3.11

It follows that

$$\sin\omega_1 = \frac{\lambda_3 h}{L_1} \tag{3.19}$$

$$\Rightarrow \frac{L_1}{h} = \frac{\lambda_3}{\sin\omega_1} \tag{3.20}$$

$$\sin\left(\frac{\pi}{3} - \omega_1\right) = \frac{\lambda_1 h}{L_1} \tag{3.21}$$

Using the rule $\sin(\alpha+\beta) = \sin\alpha\cos\beta + \sin\beta\cos\alpha$ to expand equation (3.21)

$$\frac{\sqrt{3}}{2}\cos\omega_1 - \frac{1}{2}\sin\omega_1 = \frac{\lambda_1 h}{L_1} \tag{3.22}$$

$$\Rightarrow \frac{L_1}{h} = \frac{\lambda_1}{\left(\frac{\sqrt{3}}{2}\cos\omega_1 - \frac{1}{2}\sin\omega_1\right)} \tag{3.23}$$

Equating (3.26) and (3.27) gives

$$\frac{\lambda_3}{\sin\omega_1} = \frac{\lambda_1}{\left(\frac{\sqrt{3}}{2}\cos\omega_1 - \frac{1}{2}\sin\omega_1\right)} \tag{3.24}$$

$$\Rightarrow \left(\frac{\sqrt{3}}{2}\cos\omega_1 - \frac{1}{2}\sin\omega_1\right)\lambda_3 = \lambda_1\sin\omega_1 \tag{3.25}$$

Collecting terms gives

$$\sqrt{3}\lambda_3\cos\omega_1 = (2\lambda_1 + \lambda_3)\sin\omega_1 \tag{3.26}$$

$$\Rightarrow \tan\omega_1 = \frac{\sqrt{3}\lambda_3}{(2\lambda_1 + \lambda_3)} \tag{3.27}$$

The Cartesian coordinates of point p which lie on the line A_1p are given by

$$x_p = x_{A_1} + \frac{\lambda_3 h}{\tan\omega_1} \tag{3.28}$$

and

$$y_p = y_{A_1} + \lambda_3 h \tag{3.29}$$

Substituting (3.22) into (3.23) gives

$$x_p = x_{A_1} + \frac{\lambda_3 h}{\left(\frac{\sqrt{3}\lambda_3}{(2\lambda_1 + \lambda_3)}\right)} \tag{3.30}$$

$$\Rightarrow x_p = x_{A_1} + \frac{(2\lambda_1 + \lambda_3)h}{\sqrt{3}} \tag{3.31}$$

Let the Cartesian coordinates of the vertices A_1, A_2 and A_3 be (0,0), (1,0) and ($\frac{1}{2}, \frac{\sqrt{3}}{2}$) respectively. This implies that $h = \frac{\sqrt{3}}{2}$. Substituting for h, x_{A_1} and y_{A_1} in equations (3.29) and (3.31) gives

$$x_p = \frac{(2\lambda_1 + \lambda_3)}{2} \tag{3.32}$$

$$y_p = \frac{\sqrt{3}}{2}\lambda_3 \tag{3.33}$$

Since $m_c(\{OA\})$, $m_c(\{NL\})$ and $m_c(\Theta)$ constitute a *bpa*, then the BOE_c can also be represented using simplex coordinates. As a result, both the final combined BOE_c and the individual BOE contributing to it can be shown on the same simplex plot. Two examples of this are given in Figure 3.12.

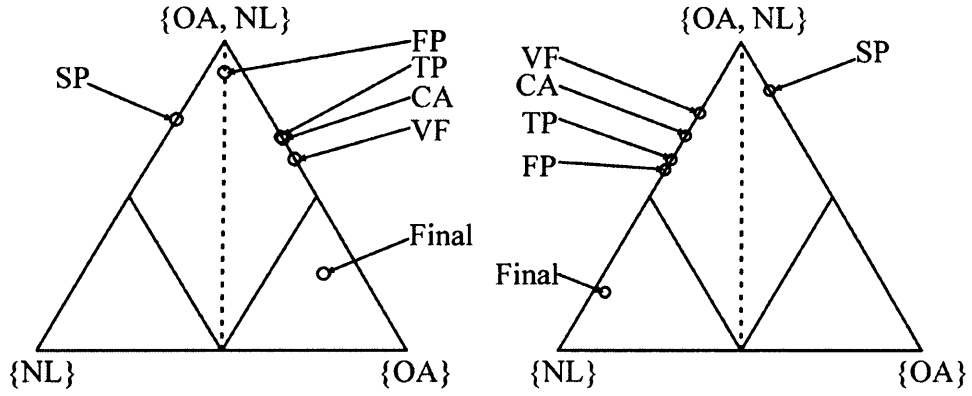


Figure 3.12 The simplex coordinate of the final combined BOE_c (Final) and the simplex coordinates of the individual BOE (SP (sagittal plane rotation), FP (frontal plane rotation), TP (transverse plane rotation), CA (cadence) and VF (vertical ground reaction force)) contributing to it can be shown on the same simplex plot (Taken from Jones *et al.*, 2003c).

The simplex plot can be divided into regions as shown in Figure 3.2c. Ignoring the $m(\Theta)$ value the simplex plot can be divided into two regions, as shown by the vertical dotted line, known here as the decision boundary. To the left of the decision boundary, $m(\{NL\}) > m(\{OA\})$ and to the right, $m(\{OA\}) > m(\{NL\})$. Therefore, it is expected that the simplex coordinates representing the final BOE of a subject classified as NL would lie to the left of the decision boundary and a subject classified as OA to the right. Further subdivision of these two regions can provide a sub classification of subjects as shown in Figure 3.2c. In region 1 of the simplex plot $m(\{NL\})$ is the dominant belief value i.e. $m(\{NL\}) > m(\{OA\}) + m(\Theta)$. This can be contrasted with region 3 in the simplex plot in which $m(\{NL\}) > m(\{OA\})$ but

$m(\{NL\}) < m(\{OA\}) + m(\Theta)$. In a similar way, in region 2 of the simplex plot $m(\{OA\})$ is the dominant factor i.e. $m(\{OA\}) > m(\{NL\}) + m(\Theta)$ whereas in region 4; $m(\{OA\}) > m(\{NL\})$ but $m(\{OA\}) < m(\{NL\}) + m(\Theta)$. Thus, a subject lying in region 2 of the simplex plot has a stronger OA classification than one lying in region 4.

3.4.5 Classification using BOE

Following the work of Safranek *et al.* (1990) the following decision rule is adopted:

i. If $m_c(\{OA\}) > m_c(\{NL\})$ then a subject is considered to have OA knee function.
ii. If $m_c(\{NL\}) > m_c(\{OA\})$ then a subject is considered to have NL knee function.

3.5 ASSIGNMENT OF VALUES TO CONTROL VARIABLES

The conversion of each input variable into a confidence factor and the subsequent construction of each BOE are dependent on a set of control variables, namely k, θ, A and B (see equations (3.1), (3.2) and (3.3), and Figure 3.2). For a set of n_i input variables, $n_i \times 4$ control variables must be evaluated. Values can be assigned to these control variables using expert knowledge or optimisation methods.

3.5.1 Assignment of values to control variables using expert knowledge

Here, assignment of values to these control variables is based on knowledge of the nature and behaviour of the input variables. The four control variables will now be discussed in turn.

Beynon (2004) suggests that an expression for k should be somehow related to the spread of the data, namely the standard deviation, σ. It is suggested that

$$k = \pm\frac{1}{\sigma} \tag{3.34}$$

where the sign depends on the association (see Figure 3.3). If there is a large spread present in the measurements of v, i.e. a large σ, then equation (3.34) will produce a small value for the confidence value increasing the range of v around $cf(v) = 0.5$, and vice-versa. Alternatively Beynon *et al.* (2002) take the k control variable as Pearson's correlation coefficient for that characteristic with the subjects' category labels (0 $\Rightarrow$ NL and 1 $\Rightarrow$ OA). The rational behind this is that the sign of the correlation coefficient gives the sign of the k value and hence the direction of the association of the characteristic with the hypothesis. The value of the correlation coefficient indicates the degree to which the characteristic can differentiate between OA and NL.

Beynon (2004) suggests that θ should not be biased towards {OA} or {NL} and thus uses the mean value, $\overline{v}$.

The A and B variables relate directly to the range of values of exact belief in {OA} and {NL}, including the level of associated uncertainty. The process of assigning values to the A and B control variables depends on knowledge of the general limits of uncertainty, $[\Theta_L, \Theta_U]$, allowed for the individual variables. Different limits can be assigned to each of the individual variables; however, in this study the same limits are assigned to each variable (following Beynon *et al.*, 2002). To use these limits to calculate A and B, reference is made back to the expressions for the belief values $m(\{OA\})$, $m(\{NL\})$ and $m(\Theta)$ (equations 3.8, 3.9 and 3.10). It can be seen from Figure 3.2b that the greatest values of uncertainty, $m(\Theta)$ are found around $cf(v) = 0.5$ and the least around $cf(v) = 0$ and 1 respectively. In the case of least uncertainty, Θ_L, using $cf(v) = 0$ in equation (3.8), then

$$m(\{OA\}) = -\frac{AB}{1-A} \tag{3.35}$$

This is always negative because A and B are positive and therefore $m(\{OA\})$ is set to zero. From equation (3.9)

$$m(\{NL\}) = B \tag{3.36}$$

By substituting these results into (3.10) it follows that

$$m(\Theta) = 1 - B \tag{3.37}$$

Substituting Θ_L for $m(\Theta)$, gives

$$B = 1 - \Theta_L \tag{3.38}$$

In the case of highest uncertainty, Θ_U, using $cf(v) = 0.5$, then from equation (3.8)

$$m(\{OA\}) = m(\{NL\}) = \frac{B(1-2A)}{1-A} \tag{3.39}$$

Subsequently, from equation (3.10) gives

$$m(\Theta) = 1 - \frac{B(1-2A)}{1-A} \tag{3.40}$$

Substituting for $m(\Theta) = \Theta_U$ and B from equation (3.38) gives

$$\Theta_U = 1 - \frac{(1-\Theta_L)(1-2A)}{1-A} \tag{3.41}$$

Rearranging gives

$$A = \frac{\Theta_U - \Theta_L}{1 + \Theta_U - 2\Theta_L} \tag{3.42}$$

3.5.2 Assignment of values to control variables using optimisation methods

In the absence of, or to eliminate the need for expert opinion and to automate the process of selecting control variables, optimisation methods can be utilised. Global

optimisation techniques such as hill climbing or simulated annealing can aid in the selection of such variables with the use of an objective function (OB) – a function that quantitatively represents a measure of a system's performance (Kirkpatrick *et al.*, 1983). In the context of this application, the OB is a measure of the level of classification of the subjects to the hypothesis {OA} or the hypothesis {NL}. The OB is dependent on the $n_i \times 4$ control variables (k, θ, A and B) and the optimisation technique seeks to minimise this OB by optimising the values of the control variables.

In reference to the simplex plot, the OB is the difference between the actual position of the subjects and the desired optimal position - that all subjects are classified correctly either as {OA} or {NL}. This implies that all subjects are positioned as close as possible to either of the base vertices in the simplex plot, ({OA} or {NL}), depending on their actual category. It follows that a suitable OB is the Euclidean distance of the mean coordinates of the two groups of subjects to their correct vertex (Beynon *et al.*, 2002)

Given that $\left(x_i^{OA}, y_i^{OA}\right)$ $i = 1,\ldots,n_{OA}$ and $\left(x_i^{NL}, y_i^{NL}\right)$ $i = 1,\ldots,n_{NL}$ are the Cartesian coordinates of the n_{OA} and n_{NL} objects in the simplex plot which are classified to the hypotheses {OA} and {NL} respectively, then OB is given as

$$OB = \frac{1}{2}\left(\sqrt{\left(\bar{x}_i^{NL} - x_{NL}\right)^2 + \left(\bar{y}_i^{NL} - y_{NL}\right)^2} + \sqrt{\left(\bar{x}_i^{OA} - x_{OA}\right)^2 + \left(\bar{y}_i^{OA} - y_{OA}\right)^2}\right) \tag{3.43}$$

where $\left(x_{OA}, y_{OA}\right)$ and $\left(x_{NL}, y_{NL}\right)$ are the simplex coordinates of the two base vertices, {OA} and {NL}, respectively.

Allowing the sides of the equilateral triangle domain of the simplex plot to be of unit length, $0 \leq OB \leq 1$. In the limit, the nearer the value is to zero the better the classification of the subjects. A value of OB near 1 infers that the mean points are near the Θ vertex or the other incorrect vertex.

Hill climbing algorithms may not be suitable for use in cases where the global minimum exists among several local minima because of the problem of local

entrapment. To illustrate the problem of local entrapment reference is made to Figure 3.13, which depicts two peaks in the Brecon Beacons national park, Pen-y-fan (point C) and Corn-Ddu (point B), the horse shoe pass (point D) and the car park in the lower valley (point A). Starting out his descent from the peak of Pen-y-fan, a hiker intent on eating his packed lunch in the car park, unexpectedly finds that he is surrounded by hill fog. Continuing downhill on his walk, he eventually reaches a point (point B) where he cannot go downwards any further. Thus, he concludes that he has reached the car park in the lower valley and sits down to eat. However, this is not the case. He has reached the horse shoe pass in between the two mountains and cannot see the lower valley because of the fog. In order to reach his true destination he must allow himself to climb upwards as well as downwards in favour of reaching the lower valley.

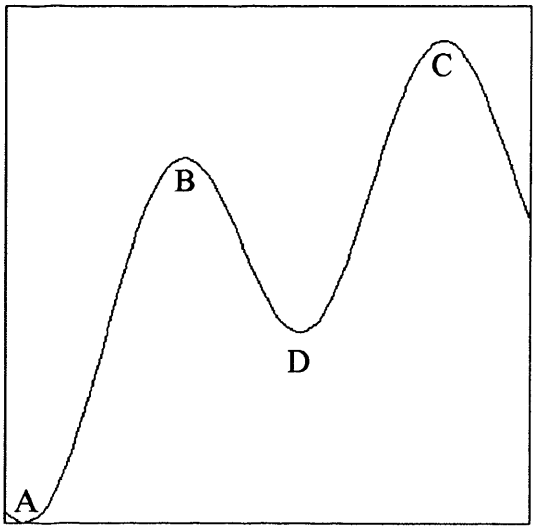


Figure 3.13 The problem of local minimum entrapment. The Brecon Beacons national park: A – car park in the lower valley, B – peak of Corn-Ddu, C – peak of Pen-y-fan and D – horse shoe pass.

This is the problem associated with the hill-climbing algorithm. Starting from an initial solution the algorithm adjusts the parameters slightly and recalculates the OB. The adjustment is accepted if, and only if, there is an improvement in the OB. As a result, the algorithm converges to the nearest local minimum at the expense of superior solutions.

Simulated annealing (SA) was introduced as an optimisation method by Kirkpatrick *et al.* (1983), based on the work of Metropolis *et al.* (1953). SA, a global optimisation technique, is able to distinguish between many local minima. Unlike the hill-climbing algorithm, it probabilistically accepts "worse" solutions and consequently escapes local minima, whilst moving towards the global minimum.

SA is analogous to the annealing process in metallurgy. This is a heating process whereby a metal is raised to an elevated temperature for an extended period and then cooled slowly, in order to relieve stresses, increase ductility and to alter the microstructure (Callister Jr., 1999). Initial temperature, time held at this temperature (soaking time), cooling rate and final temperature are crucial parameters in the annealing process:

i. If the rate of cooling is too great, internal stresses may result in warping or cracking in the material
ii. If the annealing time is too short, transformation in the microstructure may not occur
iii. If the annealing temperature is increased, the annealing process may be accelerated.

In a similar way, the performance of the SA algorithm is dependent on a number of control parameters:

i. Starting temperature
ii. Number of iterations per temperature (analogous to soaking time)
iii. Cooling rate
iv. Final temperature

The SA algorithm and these associated control parameters will now be described in more detail.

3.5.2.1 The SA algorithm

A summary of the SA algorithm is given in Figure 3.14.

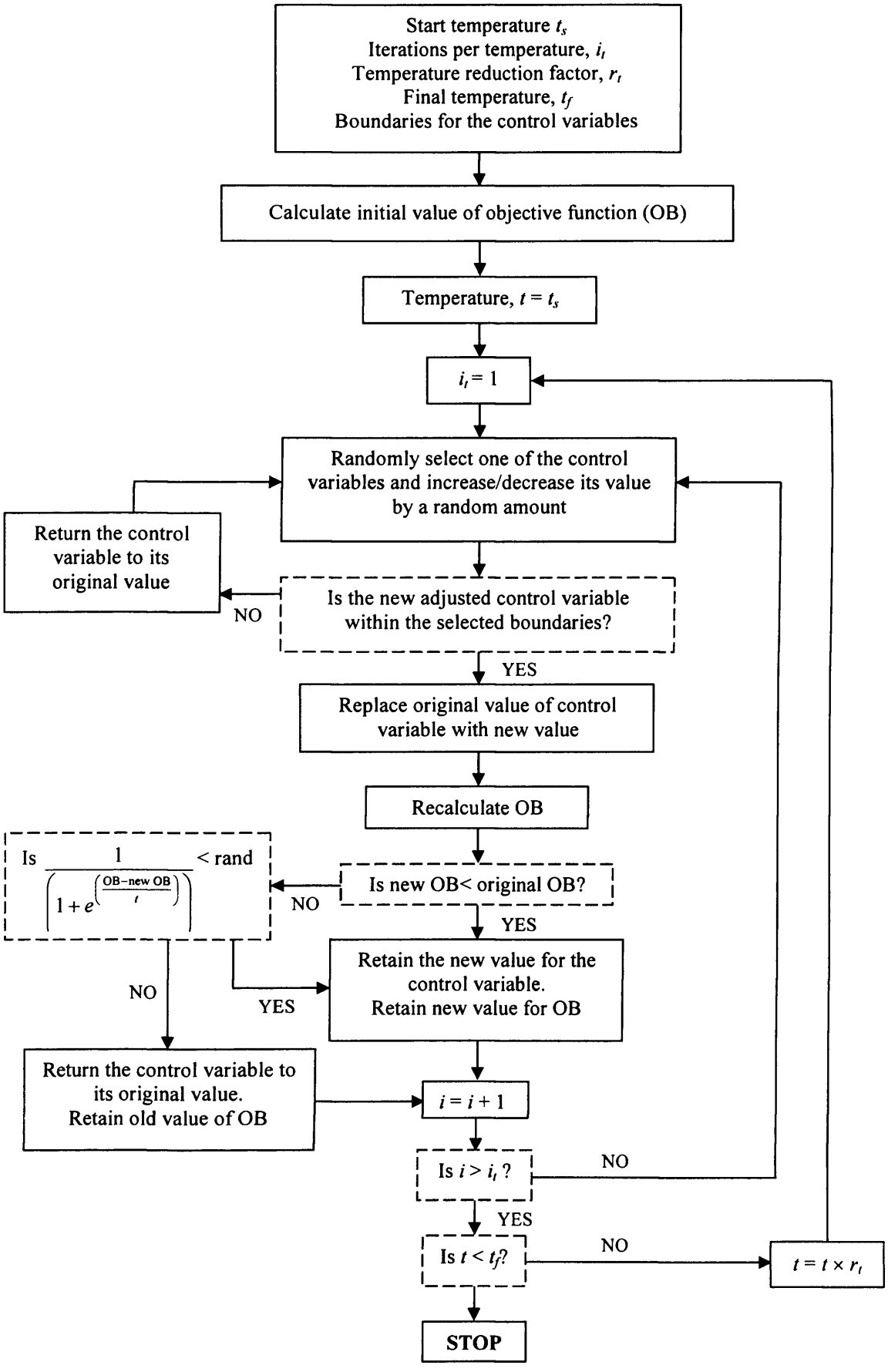


Figure 3.14 Flowchart of the simulated annealing method

i. Set the initial temperature t_s, temperature reduction factor r_t, iterations per temperature i_t and final temperature t_f (these will be discussed later in more detail).

ii. Start with an initial solution $n_i \times (k, \theta, A \text{ and } B)$. Calculate the OB as shown in Figure 3.15.

iii. Randomly choose one of the $n_i \times 4$ control variables. Adjust its value by a small amount and recalculate the OB. If the OB has decreased in value, unconditionally accept the change made to the randomly selected variable. If the change results in an increase in the OB, accept the change in the randomly selected variable according to some probability. The probability of acceptance is dependent on the current temperature t and the difference between the previous and current objective function Δf, and is given by the Metropolis criterion (Metropolis *et al.*, 1953):

$$p(t) = \frac{1}{\left(1 + e^{\frac{\Delta f}{t}}\right)} < \text{rand} \tag{3.44}$$

where rand is a random number between 0 and 1. Note that t is simply a control parameter and has no physical equivalence. Figure 3.16 shows that at higher temperatures, the probability of acceptance is around 0.5 and as a result, OB is equally likely to increase as decrease. Therefore, at high temperatures the SA performs a random walk as shown in Figure 3.17a. At lower temperatures the probability of acceptance decreases since $p(t)$ increases (Figure 3.16b). As the temperature reaches zero only better moves will be accepted. Thus, the SA algorithm behaves at low temperatures like a hill climbing algorithm (Figure 3.17b).

iv. The algorithm performs a set number of iterations (i_t) at the current temperature before the temperature is reduced according to the reduction factor r_t. This procedure continues until the stopping criterion is met.

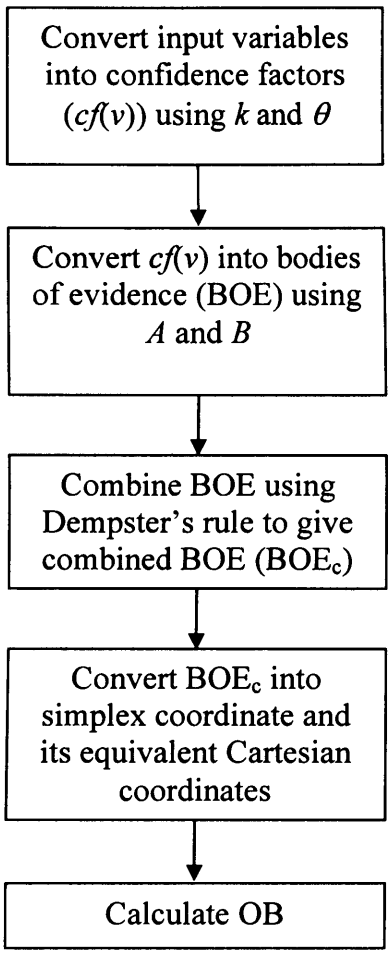


Figure 3.15 Calculation of objective function (OB)

3.5.2.2 The SA control parameters

As previously stated the performance of the SA algorithm is dependent on a number of control parameters. The significance of these parameters will now be discussed.

The temperature controls the acceptance of worse solutions (equation (3.44)) and hence allows the solution to escape local minima. If the starting temperature, t_s is not sufficiently large enough, the system will not allow for an increase in the OB. Hence, the algorithm will behave as a hill-climbing algorithm. The final solution will be near to the initial solution and the algorithm is more likely to 'get stuck' in a local minimum. Conversely, if the starting temperature is too large, the algorithm will perform a random walk and will not converge to an optimal solution.

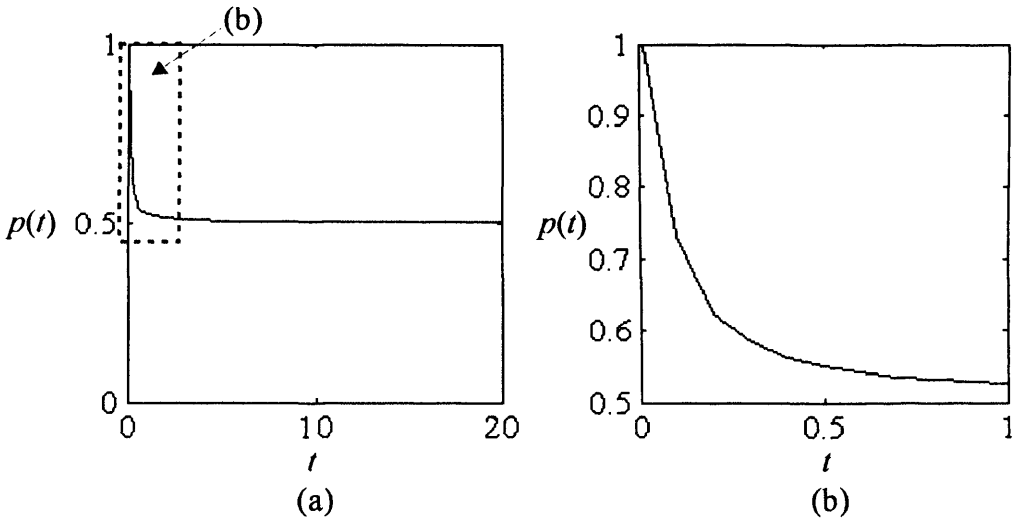


Figure 3.16 Probability distribution based on temperature for $\Delta f = -0.1$. (a) At high temperatures the $p(t) \approx 0.5$, therefore a "worse" solution is equally likely to be accepted or rejected. (b) At lower temperatures $p(t)$ approaches 1 which means that "worse" solutions are more likely to be rejected than accepted.

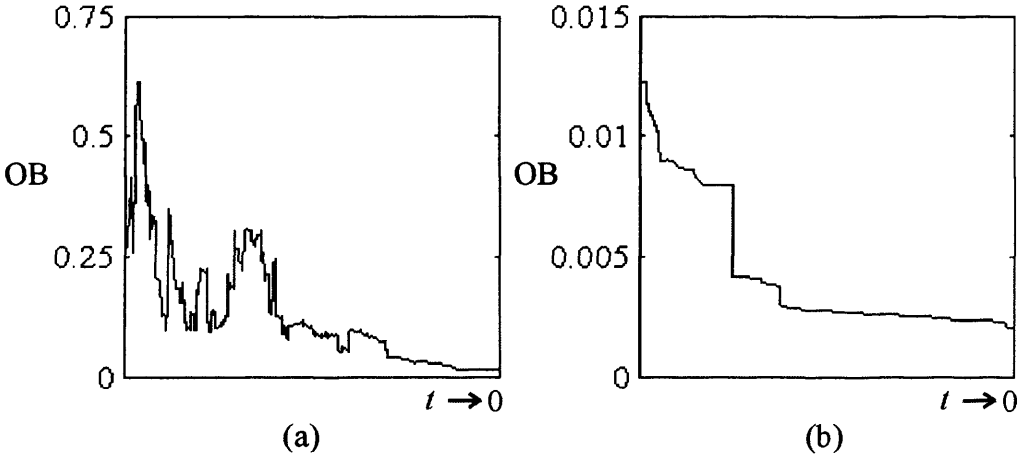


Figure 3.17 (a) At high initial temperatures, the SA performs a random walk. Therefore, the OB is allowed to increase as well as decrease. (b) As the temperature t decreases, the SA performs like a hill-climbing algorithm, and the OB is only allowed to decrease.

Duda *et al.* (2001) comment that the initial temperature must be sufficiently large so that all configurations have equal probability of acceptance. Nolle *et al.* (2002) use an initial starting temperature based on knowledge of OB.

In the same way as Duda *et al.* (2001), Nolle *et al.* (2002) suggest a value of 0.5 for the initial transition probability if the difference in fitness, Δf, is small. Using this criterion of $p(t_s) \approx 0.5$, gives:

$$\frac{1}{1+e^{\Delta f / t_s}} = 0.49$$

$$\Rightarrow t_s \approx 25 \times \Delta f \tag{3.45}$$

The time spent at each temperature is usually related to the size of the neighbourhood or possibly the size of the solution space (Dowsland, 1995).

Two different types of temperature reduction are most commonly used (Dowsland, 1995). The first method invokes a geometric reduction function

$$\alpha(t) = at \tag{3.46}$$

where $a<1$.

The second method performs one iteration at each temperature but reduces the temperature very slowly using the formula

$$\alpha(t) = \frac{t}{(1+\beta t)} \tag{3.47}$$

where β is a small value. Other methods of temperature cooling are summarised by Dowsland (1995).

Theoretically, the temperature should be allowed to reach zero before the algorithm is stopped. However, this results in unfeasibly large computational times. Furthermore,

before reaching a temperature of zero, at relatively small temperatures, the probability of acceptance of uphill moves will be indistinguishable from zero. Thus, it is not necessary to run the algorithm until a temperature of zero is reached. To reduce the strictness of the final temperature equalling zero, one or more stopping constraints may be employed. Dowsland (1995) suggests that the simplest method is to pre-specify the number of iterations and stop when this number has been completed. However, this final number of iterations must correspond to a sufficiently low temperature to ensure convergence. A second stopping constraint is to run the algorithm until a final temperature is reached (Dowsland, 1995).

The application of SA to the classification of OA and NL subjects is given in Chapter 4, section 4.2.

3.6 EVALUATION

The final stage of classifier design is performance evaluation (Duda *et al.*, 2001). Siedlecki and Sklansky (1989) comment that the only legitimate way of evaluating the performance of a classifier is through examination of its error or misclassification rate:

$$\text{error rate} = \frac{\text{number of misclassifications}}{\text{number of cases}} \tag{3.48}$$

This evaluation method is also the most commonly used (Weiss and Kulikowski, 1991). For a discussion of other evaluation methods the reader is directed to Dash and Liu (1997). The true error rate of a classifier is defined as the error rate on an infinite number of new cases (Weiss and Kulikowski, 1991). However, in reality the number of cases is finite and the true error rate must be approximated.

The most obvious way of approximating the true error rate is to use the resubstitution (or apparent or reclassification) error rate. In this instance the classifier is trained using a set of cases and the error rate is calculated using the same cases that were used to design the classifier (Weiss and Kulikowski, 1991). However, the resubstitution error is known to be an optimistically biased estimate of the true error rate (Raudys

and Jain, 1991; Toussaint, 1974) and consequently should only be used when the number of training cases is large.

Different methods exist for obtaining less biased estimates of the true error rate. These methods involve partitioning the cases into a training group and a testing group. The simplest of these methods is known as the hold-out method (see Toussaint, 1974). Here a set number of cases are assigned to the training group and the remainder to the testing group. The classifier is trained on the training group and then the hold-out error is calculated using the testing group. For large testing group sizes the hold-out error approaches the true error. However, for smaller sample sizes the hold-out method is a pessimistically biased estimate of the true error rate (Toussaint, 1974) and is highly dependent on the way in which the samples are partitioned into a training and testing group (Toussaint, 1974). Furthermore, the hold-out method makes inefficient use of the data (Toussaint, 1974; Raudys and Jain, 1991) especially when there are limited number of training and testing cases.

Instead of estimating the error rate using a single testing and training group (the hold-out method), cross-validation or resampling methods that use multiple testing and training groups can be used (Weiss and Kulikowski, 1991). Using these methods a number of classifiers are created. The data (N samples) is partitioned into a training group (n samples) and a testing group ($N-n$ samples) and the classifier is trained on the training data and tested on the testing data as before. This process is repeated using different training groups of size n. The cross-validation error rate is then defined as the average testing group error rate for all of the classifiers.

A special case of cross-validation is the leave-one-out method. Here, the classifier is trained on the (N–1) training cases and tested on the remaining one test case. This process is repeated N times. The leave-one-out error rate is then defined as the average test case error rate. This method overcomes the issue of inefficient use of the data since every case is used in testing and each time every case save one is employed as a training case (Weiss and Kulikowski, 1991). Additionally, and perhaps more importantly the leave-one-out error rate is nearly an unbiased estimate of the true error rate even for small sample sizes (Raudys and Jain, 1991; Weiss and Kulikowski,

1990). However, there are disadvantages to the leave-one-out method. Firstly the leave-one-out error rate has a large associated variance, especially for small sample sizes (Toussaint, 1974; Weiss and Kulikowski, 1991). Secondly it can be computationally intensive to calculate the leave-one-out error rate for large sample sizes. Following the recommendation of Weiss and Kulikowski (1991) the leave-one-out method is used to estimate the true error rate of the classifier in the studies carried out as part of this thesis. Using a sample of 42 subjects, the DS classifier is trained on 41 of the training cases and tested on the remaining one test case. The process is repeated 42 times and the leave-one-out error is calculated as the average of the average error rate for all of the left-out samples. The out-of-sample accuracy is defined as 100 minus the leave-one-out error.

This chapter has introduced and described the new classification method in detail. Conclusions based on this chapter are given in Chapter 8.

CHAPTER 4

DEMPSTER-SHAFER CONTROL PARAMETERS

This chapter reports the results obtained from a series of tests that were conducted to investigate the influence of the Dempster-Shafer (DS) control parameters on the accuracy and stability of the DS classifier. This study was necessary to investigate the validity of the DS classifier. Values can be assigned to the DS control parameters using one of two methods as outlined in section 3.5. The first method, a non-optimisation approach uses expert opinion and knowledge of the input variables to assign values to the DS control parameters. The second method, an optimisation approach, uses simulated annealing to assign values to the DS control parameters. The influence of the DS control parameters using both methods is investigated in sections 4.1 and 4.2 respectively. A summary based on the work is given in section 4.3.

4.1 NON-OPTIMISATION METHOD

4.1.1 Sensitivity of the DS control parameters to changes in the subject population and in the uncertainty limits

Using the non-optimisation method values are assigned to the DS control parameters, k, θ, A and B, using a set of equations developed in section 3.5.1. Two different definitions of k (k_c and k_s) were established. For a given input variable v_i, k_c is given as the Pearson's correlation coefficient for v_i with the subjects' category label (0 $\Rightarrow$ NL and 1 $\Rightarrow$ OA), and k_s is given as

$$k_s = \pm\frac{1}{\sigma} \tag{4.1}$$

where σ is the standard deviation about the mean of v_i and the sign of k_s is determined from the sign of Pearson's correlation coefficient. The absolute value of k dictates the range of measurement values for which the confidence function for $cf(v)$ is near 0.5. A positive value of k implies that a large v_i measurement offers more support to {OA}

($cf(v) \rightarrow 1$), whilst a small v_i measurement offers more support to {NL} ($cf(v) \rightarrow 0$). Conversely, a negative k implies that a large v_i measurement offers more support to {NL} whilst a small v_i measurement offers more support to {OA}.

For v_i, θ is defined as the population mean, $\overline{v_i}$. θ determines the value of v_i for which $cf(v) = 0.5$.

The DS control parameters A and B are dependent on the limits of uncertainty $[\Theta_L, \Theta_U]$ and are defined as

$$A = \frac{\Theta_U - \Theta_L}{1 + \Theta_U - 2\Theta_L} \tag{4.2}$$

$$B = 1 - \Theta_L \tag{4.3}$$

The DS control parameters k and θ are dependent on the subject population (the OA and NL combined sample of 42 subjects). An investigation was conducted to determine the sensitivity of k and θ to changes in the subject population. Using a leave-one-out (LOO) approach, the values of k and θ for each input variable were calculated 42 times, using the 42 combinations of 41 subjects. This was performed for both definitions of k, namely k_c and k_s.

The mean values of k_c and k_s for v_i (i = 1:18) are given in Appendix B (Table B.1) and depicted in Figure 4.1.

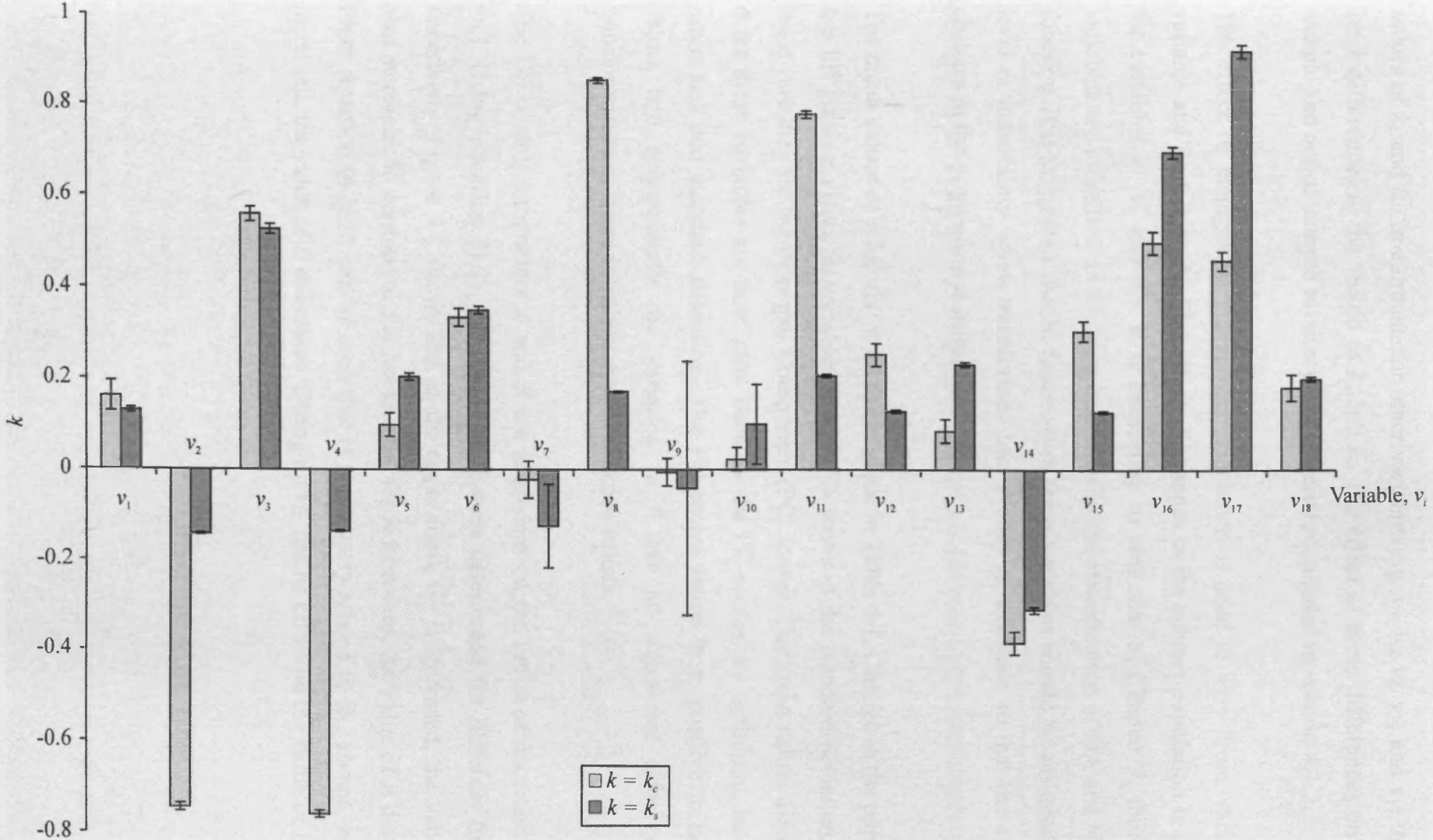


Figure 4.1 The Sensitivity of k to changes in the population for each input variable, v_i. The average values of k_c *and* k_s are given with error bars indicating ± one standard deviation.

For some variables, e.g. v_1, v_6 and v_{18} there are only small differences between the values of k_c and k_s. In contrast, for other variables e.g. v_2, v_4, v_8, v_{11} and v_{17} there are large differences in the values of k_c and k_s. The effect of these differences on the in-sample and out-of-sample accuracy and OB are investigated in section 4.1.2.

The effect of changes in the subject population is seen to vary from variable to variable and from k_c to k_s. The effect of changes in the subject population is greatest for variables v_7, v_9 and v_{10}. It is interesting to note that in Chapter 7, these three variables are identified as the least important in the classification of OA and NL knee function. It is suggested that in future work these variables should be assigned a high level of uncertainty when transformed into a body of evidence so that the effect of changes in the population is minimised.

The mean values of θ for v_i (i = 1:18) are given in Table 4.1. Change in the population has the greatest effect on variables v_4 to v_{15} in terms of the standard deviation. All of these variables are the Principal Component (PC) scores. The mean values assigned to θ for these variables are near zero, because the PC scores, by definition, have zero mean and unit standard deviation. The PC scores range from positive to negative values and consequently the variation in θ has no significant effect on the transformation of these variables into confidence values.

The DS control parameters A and B are dependent on the limits of uncertainty [Θ_L, Θ_U]. Using equation (4.2), the values of A were determined for different limits of uncertainty. Figure 4.2 shows that as the upper limit, Θ_U is increased, the value of A also increases. In contrast as the lower limit, Θ_L is increased, the value of A decreases. From equation (4.3), it can be seen that B is inversely related to Θ_L. Hence, as Θ_L is increased, the value of B decreases. Changing Θ_U has no effect on the value of B.

Table 4.1 The sensitivity of θ to changes in population for each input variable, v_i

Variable	Mean	Standard Deviation
v_1	27.2788	0.1872
v_2	49.0004	0.1712
v_3	60.8986	0.0466
v_4	-1.90E-08	0.1791
v_5	9.07E-09	0.1189
v_6	2.05E-09	0.0697
v_7	-5.10E-09	0.1609
v_8	-9.70E-09	0.1428
v_9	-1.00E-08	0.0879
v_{10}	-4.50E-09	0.1842
v_{11}	-9.50E-09	0.1182
v_{12}	1.26E-08	0.1890
v_{13}	-1.30E-08	0.1046
v_{14}	-3.40E-09	0.0789
v_{15}	3.21E-08	0.1896
v_{16}	10.4500	0.0353
v_{17}	12.0595	0.0267
v_{18}	42.2691	0.1218

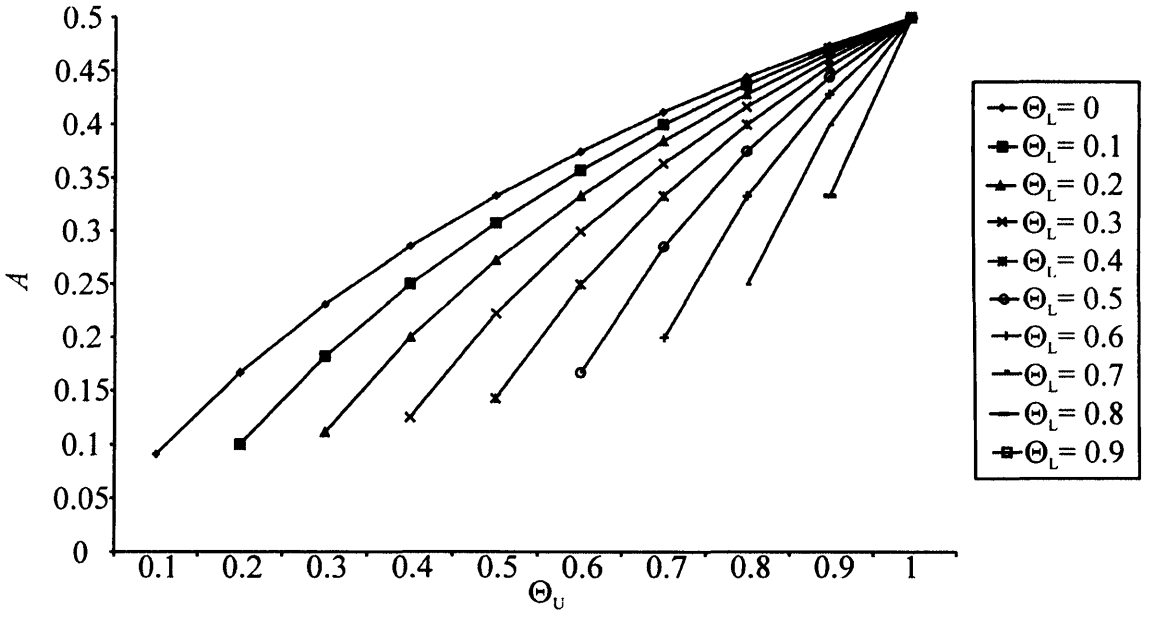


Figure 4.2 The effect of changing Θ_L and Θ_U on the DS control parameter A.

4.1.2 The effect of changing the DS control parameters on DS classifier performance

A series of tests were conducted to examine the effect of using the two different definitions of the DS control parameter k, namely k_c and k_s, and of changing the uncertainty limits $[\Theta_L, \Theta_U]$, on the performance of the DS classifier. Using a LOO approach the DS classifier accuracy and objective function (OB) were calculated. As stated in Chapter 3 (section 3.5.2), the OB is a measure of the level of certainty in the classification of subjects to their assigned class, where a value of OB close to zero implies a more robust classification than a value near to one. This was carried out for both definitions of k and for the different uncertainty limits. A description of the tests is given in Table 4.2. The test number (1 to 10) identifies the value of Θ_U; the subtest (I or II) specifies the definition of k; the part (a to j) indicates the value of Θ_L. The number of parts is dependent on the value of Θ_U, since Θ_L must be less than Θ_U. As an example, Table 4.2 shows that in test 3IIf the DS classifier is trained with k_s and uncertainty boundaries [0.5, 0.8].

Table 4.2 Description of tests to determine the effect of changing the definition of k and the uncertainty boundaries $[\Theta_L, \Theta_U]$ on the DS classifier accuracy and OB.

Test	Θ_U
1	1
2	0.9
3	0.8
4	0.7
5	0.6
6	0.5
7	0.4
8	0.3
9	0.2
10	0.1

Subtest	k definition
I	k_c
II	k_s

Part	Θ_L
a	0
b	0.1
c	0.2
d	0.3
e	0.4
f	0.5
g	0.6
h	0.7
i	0.8
j	0.9

4.1.2.1 Test 1

Test 1 investigates the effect of changing the definition of k and the value of Θ_L on the accuracy and OB of the DS classifier when $\Theta_U = 1$. The average values of the in- and out-of-sample accuracy and the in- and out-of-sample OB are tabulated in Appendix B (Table B.2) and depicted in Figures 4.3 and 4.4.

The positioning of the simplex coordinates of the 42 out-of-samples subjects in the simplex plot is given in Figure 4.5 for the DS classifiers with the highest in-sample accuracies i.e. the DS classifiers from tests 1Ib to 1Ij respectively. These figures were chosen to illustrate the effect of increasing Θ_L on the positioning of the out-of-sample subjects in the simplex plot domain. Furthermore, although obtaining high classification accuracy is of first importance, these figures highlight that the simplex plot helps to identify the practical and clinical relevance of the DS classifier. The simplex plots for test 1Ia and tests 1IIa to 1IIj are given in Appendix B (Figures B.1 to B.11 respectively).

For the DS classifiers trained with k_c (subtest I) the in-sample accuracy increases from 97.56% to a maximum value of 97.62% as Θ_L is increased from 0 to 0.1. For all other values of Θ_L, the in-sample accuracy remains constant at this maximum value. For the DS classifiers trained with k_s (subtest II) there is no set pattern in the behaviour of the in-sample accuracy as Θ_L is increased from 0 to 0.9.

For k_c, there is no change in the value of the out-of-sample accuracy as Θ_L is increased from 0 to 0.9. At all values of Θ_L the DS classifier has an out-of-sample accuracy of 97.62%. For k_s the out-of-sample accuracy increases from 83.33% to 88.10% as Θ_L is increased from 0 to 0.1. The out-of-sample accuracy remains at this value as Θ_L is increased to 0.4. As Θ_L is increased from 0.4 to 0.5, the out-of-sample accuracy falls to a value of 85.71%. Finally, the out-of-sample accuracy returns to a value of 88.10% and remains at this value as Θ_L is increased further.

For both k_c and k_s, the in-sample OB shows an increasing trend as Θ_L is increased from 0 to 0.9. As for the in-sample accuracy, the out-of-sample OB shows an increasing trend as Θ_L is increased from 0 to 0.9 for both k_c and k_s.

The DS classifier trained with k_c has higher in- and out-of-sample accuracies and lower in- and out-of-sample OB than that trained with k_s.

In each of the simplex plots in Figure 4.5 (Figures 4.5a to 4.5i), 41 of the 42 out-of-sample subjects have been correctly classified. The simplex coordinates of all 20 OA subjects lie nearer to the {OA} vertex than the {NL} vertex (on the right hand side of the decision boundary), since for these subjects, $m_c(\{OA\}) > m_c(\{NL\})$. 21 of the 22 NL subjects lie nearer to the {NL} vertex than to the {OA} vertex (on the left hand side of the decision boundary). The misclassified NL subject lies on the incorrect side of the decision boundary. For this subject, $m_c(\{OA\}) > m_c(\{NL\})$. It is significant that on review of the NL cohort, it has been noted that the misclassified subject has subsequently been diagnosed with knee problems (osteoporosis), which were not noted at the time of measurement.

In Figure 4.5a the simplex coordinates of all subjects are positioned along the base opposite the Θ vertex. This implies that the level of uncertainty associated with the

combined body of evidence is near zero i.e. $m_c(\Theta) \approx 0$. Consequently, there is a very high level of certainty in the final classification of all subjects, including the misclassified NL subject. All of the simplex coordinates of the correctly classified NL subjects lie very near to the NL vertex of the simplex plot. This implies that for each of these subjects $m_c(\{OA\}) \approx 0$ and $m_c(\{NL\}) \approx 1$. This positioning of the simplex coordinates of the 21 correctly classified NL subjects implies that each of these subjects has an almost identical level of NL classification, suggesting a limited amount of variation in the NL population. This does not correspond to clinical findings, which suggest that there is a high level of variability in the NL subject population. The author notes form experience of working on the knee clinical trial, that there was a great deal of variability in the gait patterns of subjects with NL knee function.

As Θ_L is increased, the distance of the simplex coordinates of all subjects from the Θ vertex decreases. This is a result of the increase in $m_c(\Theta)$, i.e., the increase in the level of uncertainty in the final classification of all subjects. For example, as Θ_L is increased from 0.1 to 0.6 (comparing Figures 4.5a and 4.5f), the simplex coordinates of all 42 subjects no longer lie on the base opposite the Θ vertex. Comparison of Figures 4.5a and 4.5f also shows that the distance of the simplex coordinate of the misclassified NL subject from the edge opposite the NL vertex has increased as Θ_L is increased from 0 to 0.6. This suggests that $m_c(\{NL\})$ has increased and that the subject now has an increased level of NL classification but is still closer to the OA vertex. In Figure 4.5f, the correctly classified subjects all lie within the dominant classification regions of the simplex plot as in Figure 4.5a. However, in Figure 4.5f, the simplex coordinates of these subjects do not lie on their respective vertices as in Figure 4.5a, but the simplex coordinates are spread out within the dominant regions. This suggests a greater variability of NL subjects within the NL sample and of the OA subjects within the OA sample as would be expected clinically.

In Figures 4.5a to 4.5f (i.e. when $\Theta_L \leq 0.6$), the simplex coordinates of the 42 out-of-sample subjects lie within the dominant areas of the simplex plot. However, increasing Θ_L to 0.7 results in the movement of the simplex coordinates of four OA subjects from the dominant OA classification region to the non-dominant OA classification region of the simplex plot as shown in Figure 4.5g. Increasing Θ_L from

0.7 to 0.8, results five more subjects moving from dominant to non-dominant regions of the simplex plot. Figure 4.5h shows that this includes the misclassified NL subject, which has moved from the dominant OA to the non-dominant OA region of the simplex plot. As Θ_L is increased further to 0.9, the simplex coordinates of the majority of subjects move to within the non-dominant regions of the simplex plot. Despite having a high level of classification accuracy, the lack of dominant OA and NL classifications signify that the classifier is becoming clinically impractical.

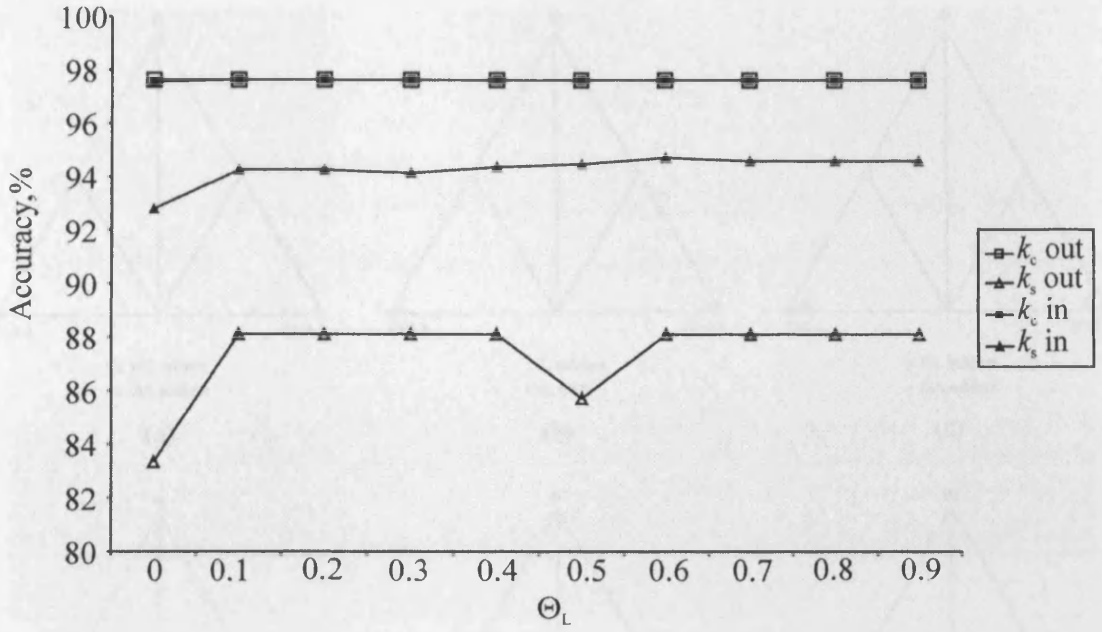


Figure 4.3 Test 1 in-sample (in) and out-of-sample (out) accuracy results showing the effect of changing Θ_L when $\Theta_U = 1$. The results for the two different definitions of k, k_c and k_s, are presented.

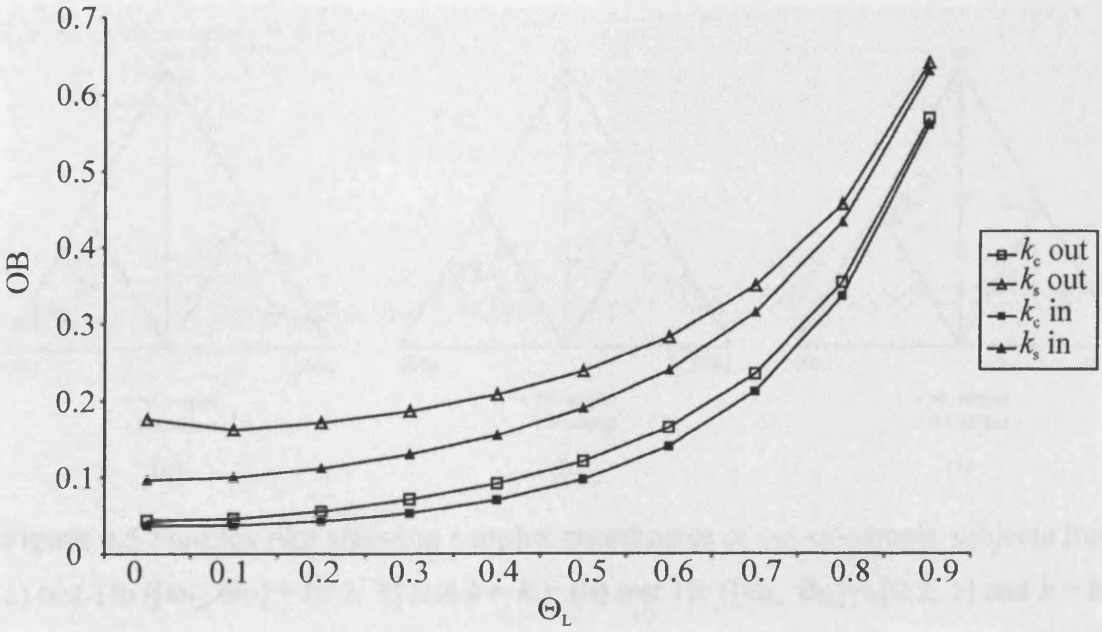


Figure 4.4 Test 1 in-sample (in) and out-of-sample (out) OB results showing the effect of changing Θ_L when $\Theta_U = 1$. The results for the two different definitions of k, k_c and k_s, are presented.

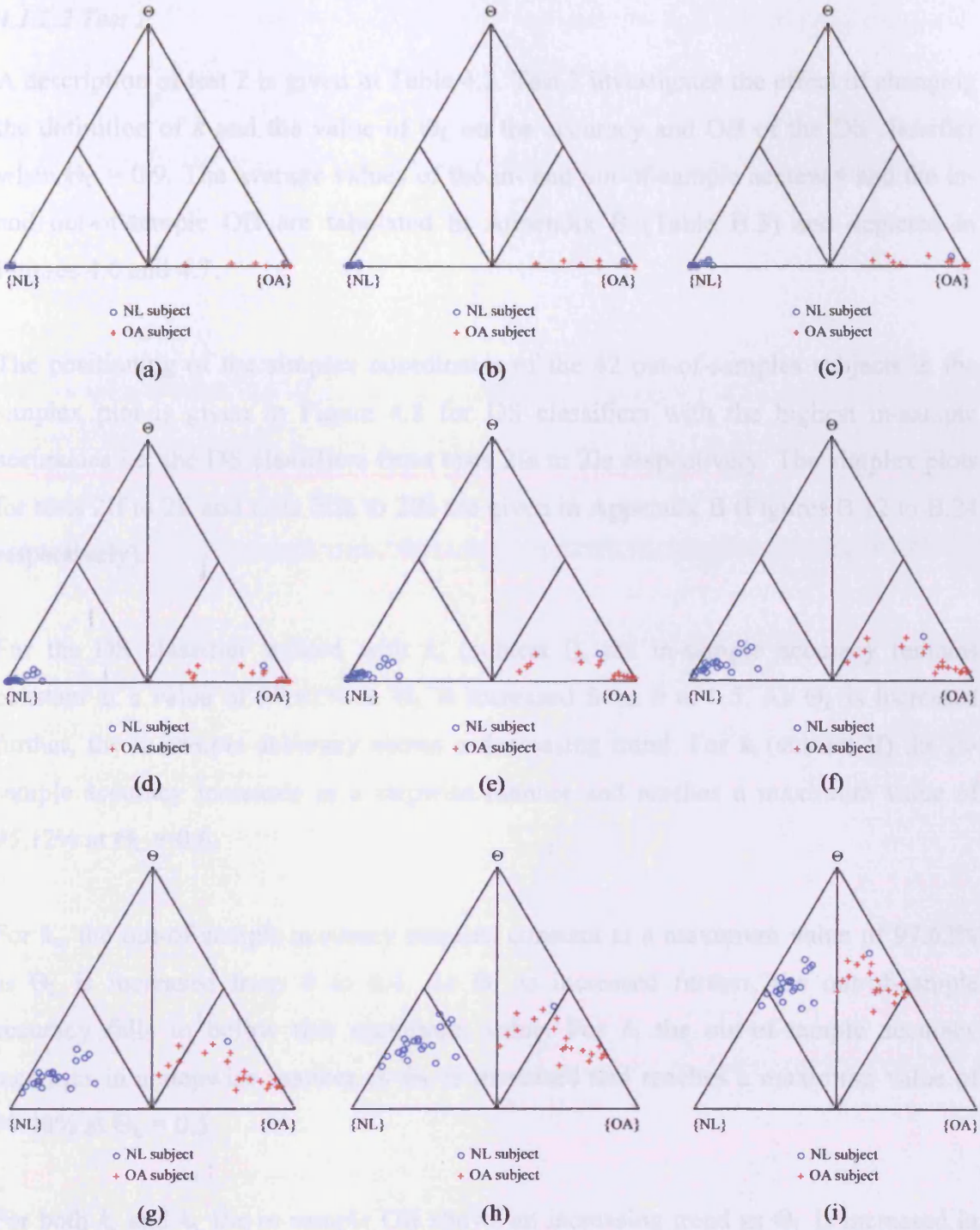


Figure 4.5 Simplex plot showing simplex coordinates of out-of-sample subjects from (a) test 1Ib ($[\Theta_L, \Theta_U] = [0.1, 1]$ and $k = k_c$); (b) test 1Ic ($[\Theta_L, \Theta_U] = [0.2, 1]$ and $k = k_c$); (c) test 1Id ($[\Theta_L, \Theta_U] = [0.3, 1]$ and $k = k_c$); (d) test 1Ie ($[\Theta_L, \Theta_U] = [0.4, 1]$ and $k = k_c$); (e) test 1If ($[\Theta_L, \Theta_U] = [0.5, 1]$ and $k = k_c$); (f) test 1Ig ($[\Theta_L, \Theta_U] = [0.6, 1]$ and $k = k_c$); (g) test 1Ih ($[\Theta_L, \Theta_U] = [0.7, 1]$ and $k = k_c$); (h) test 1Ii ($[\Theta_L, \Theta_U] = [0.8, 1]$ and $k = k_c$); (i) test 1Ij ($[\Theta_L, \Theta_U] = [0.9, 1]$ and $k = k_c$).

4.1.2.2 Test 2

A description of test 2 is given in Table 4.2. Test 2 investigates the effect of changing the definition of k and the value of Θ_L on the accuracy and OB of the DS classifier when $\Theta_U = 0.9$. The average values of the in- and out-of-sample accuracy and the in- and out-of-sample OB are tabulated in Appendix B (Table B.3) and depicted in Figures 4.6 and 4.7.

The positioning of the simplex coordinates of the 42 out-of-samples subjects in the simplex plot is given in Figure 4.8 for DS classifiers with the highest in-sample accuracies i.e. the DS classifiers from tests 2Ia to 2Ie respectively. The simplex plots for tests 2If to 2Ii and tests 2IIa to 2IIi are given in Appendix B (Figures B.12 to B.24 respectively).

For the DS classifier trained with k_c (subtest I), the in-sample accuracy remains constant at a value of 97.62% as Θ_L is increased from 0 to 0.5. As Θ_L is increased further, the in-sample accuracy shows a decreasing trend. For k_s (subtest II) the in-sample accuracy increases in a stepwise manner and reaches a maximum value of 95.12% at $\Theta_L = 0.6$.

For k_c, the out-of-sample accuracy remains constant at a maximum value of 97.62% as Θ_L is increased from 0 to 0.4. As Θ_L is increased further, the out-of-sample accuracy falls to below this maximum value. For k_s the out-of-sample accuracy increases in a stepwise manner as Θ_L is increased and reaches a maximum value of 90.48% at $\Theta_L = 0.5$.

For both k_c and k_s, the in-sample OB shows an increasing trend as Θ_L is increased in increments of 0.1 from 0 to 0.8.

For k_c, the out-of-sample OB shows an increasing trend as Θ_L is increased from 0 to 0.8. For k_s, as Θ_L is increased from 0 to 0.1, the out-of-sample OB decreases. As Θ_L is increased further, the out-of-sample OB increases.

The DS classifier trained with k_c produces a higher in- and out-of-sample accuracy and a lower in- and out-of-sample OB than that trained with k_s.

41 of the 42 out-of-sample subjects have been correctly classified. The misclassified NL subject lies on the incorrect side of the decision boundary. For this subject $m_c(\{OA\}) > m_c(\{NL\})$. The simplex coordinates of all out-of-sample subjects lie within dominant regions of the simplex plot.

The simplex coordinates of all subjects are positioned along or very near to the base opposite the Θ vertex. This implies that the level of uncertainty associated with the combined body of evidence is near zero i.e. $m_c(\Theta) \approx 0$. Consequently, there is a very high level of certainty in the final classification of all subjects, including the misclassified NL subject. The positioning of the simplex coordinates of the 21 correctly classified NL subjects suggests that each of these subjects has an almost identical level of NL classification, implying a limited amount of variation in the NL population in terms of knee function. This does not correspond to clinical findings, which suggest that there is a high level of variability in the NL subject population.

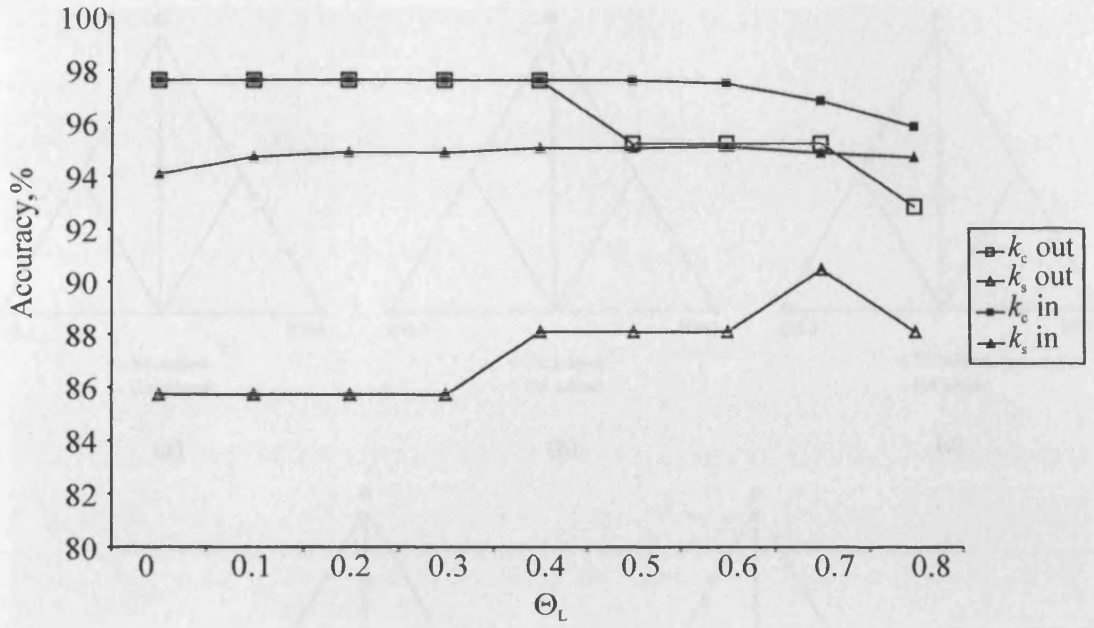


Figure 4.6 Test 2 in-sample (in) and out-of-sample (out) accuracy results showing the effect of changing Θ_L when $\Theta_U = 0.9$. The results for the two different definitions of k, k_c and k_s are presented.

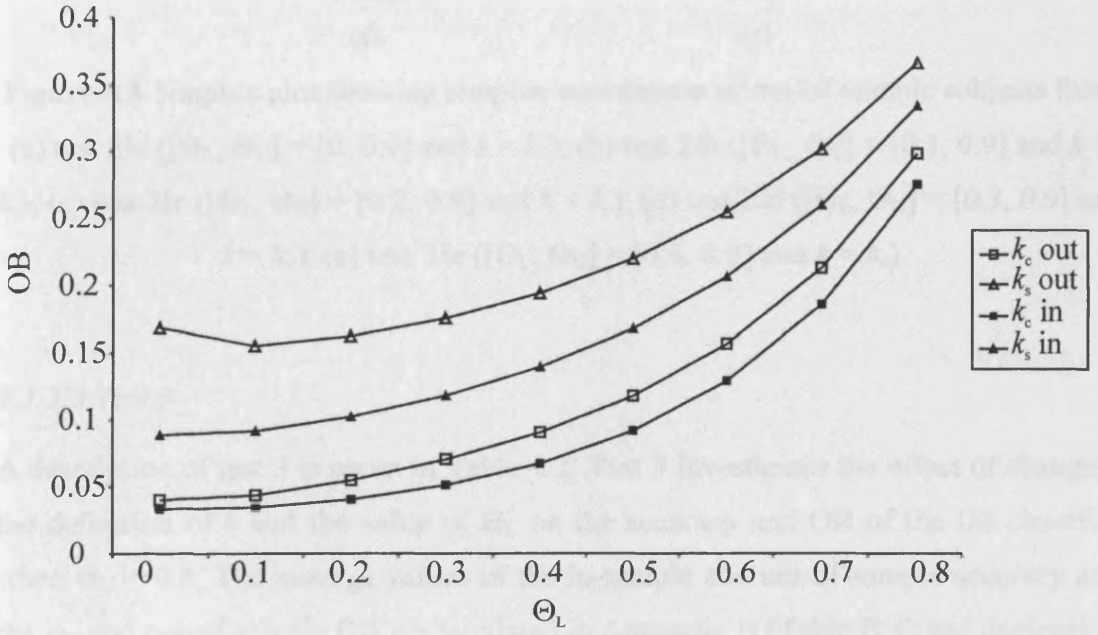


Figure 4.7 Test 2 in-sample (in) and out-of-sample (out) OB results showing the effect of changing Θ_L when $\Theta_U = 0.9$. The results for the two different definitions of k, k_c and k_s are presented.

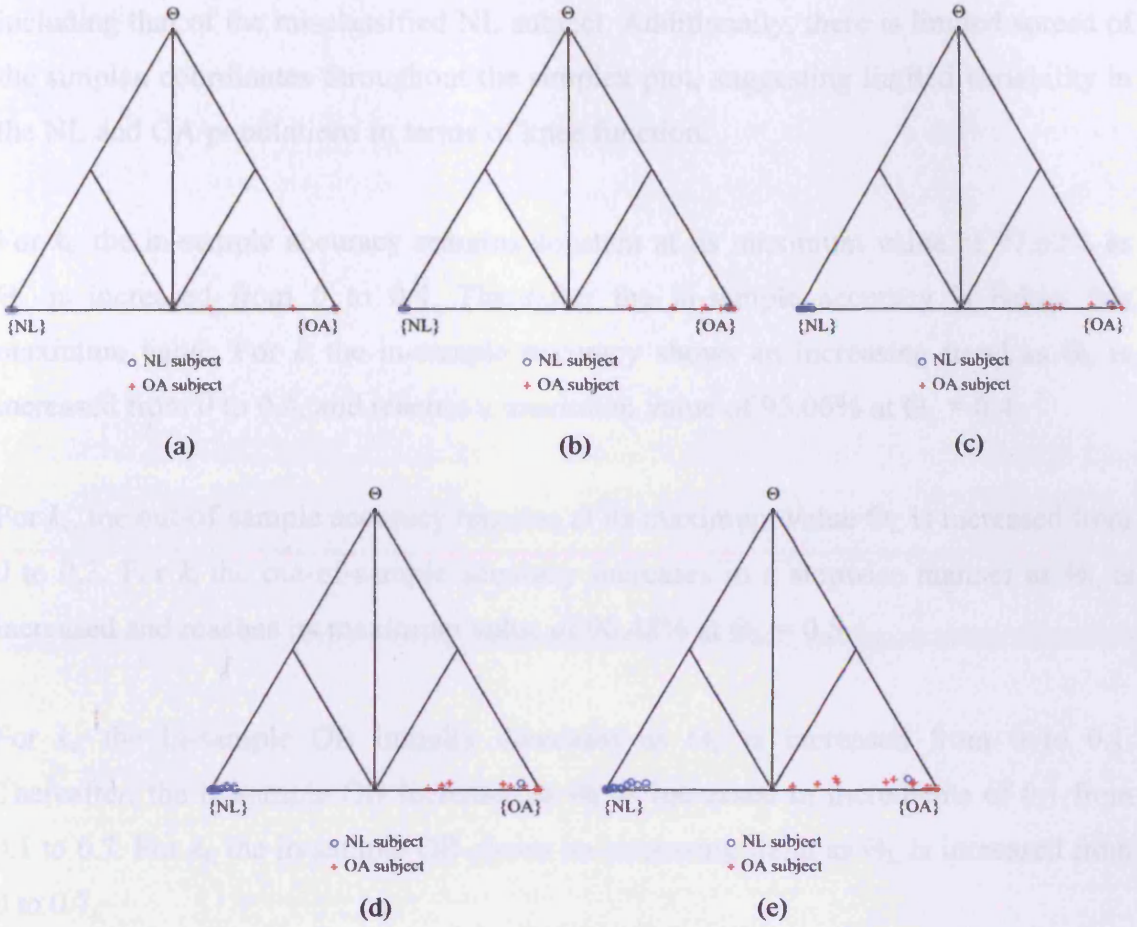


Figure 4.8 Simplex plot showing simplex coordinates of out-of-sample subjects from (a) test 2Ia ($[\Theta_L, \Theta_U] = [0, 0.9]$ and $k = k_c$); (b) test 2Ib ($[\Theta_L, \Theta_U] = [0.1, 0.9]$ and $k = k_c$); (c) test 2Ic ($[\Theta_L, \Theta_U] = [0.2, 0.9]$ and $k = k_c$); (d) test 2Id ($[\Theta_L, \Theta_U] = [0.3, 0.9]$ and $k = k_c$); (e) test 2Ie ($[\Theta_L, \Theta_U] = [0.4, 0.9]$ and $k = k_c$)

4.1.2.3 Test 3

A description of test 3 is given in Table 4.2. Test 3 investigates the effect of changing the definition of k and the value of Θ_L on the accuracy and OB of the DS classifier when $\Theta_U = 0.8$. The average values of the in-sample and out-of-sample accuracy and the in- and out-of-sample OB are tabulated in Appendix B (Table B.4) and depicted in Figures 4.9 to 4.10. The positioning of the simplex coordinates of the 42 out-of-samples subjects in the simplex plot is given in Appendix B (Figures B.25 to B.40 respectively). Briefly, for the DS classifiers with the maximum out-of-sample accuracy of 97.62%, all subjects lie within the dominant regions of the simplex plot, positioned along or very near to the base opposite the Θ vertex. This implies a high level of certainty in the classification of each subject to their respective class,

including that of the misclassified NL subject. Additionally, there is limited spread of the simplex coordinates throughout the simplex plot, suggesting limited variability in the NL and OA populations in terms of knee function.

For k_c, the in-sample accuracy remains constant at its maximum value of 97.62% as Θ_L is increased from 0 to 0.4. Thereafter the in-sample accuracy is below this maximum value. For k_s the in-sample accuracy shows an increasing trend as Θ_L is increased from 0 to 0.4, and reaches a maximum value of 95.06% at $\Theta_L = 0.4$.

For k_c, the out-of-sample accuracy remains at its maximum value Θ_L is increased from 0 to 0.3. For k_s the out-of-sample accuracy increases in a stepwise manner as Θ_L is increased and reaches its maximum value of 90.48% at $\Theta_L = 0.5$.

For k_c, the in-sample OB initially decreases as Θ_L is increased from 0 to 0.1. Thereafter, the in-sample OB increases as Θ_L is increased in increments of 0.1 from 0.1 to 0.7. For k_s, the in-sample OB shows an increasing trend as Θ_L is increased from 0 to 0.7.

For k_c, the out-of-sample OB shows an increasing trend as Θ_L is increased from 0 to 0.8. For k_s, as Θ_L is increased from 0 to 0.1, the out-of-sample OB decreases. As Θ_L is increased further, the out-of-sample OB increases.

The DS classifier that is trained with k_c consistently produces a higher in- and out-of-sample accuracy and a lower in- and out-of-sample OB than that trained with k_s.

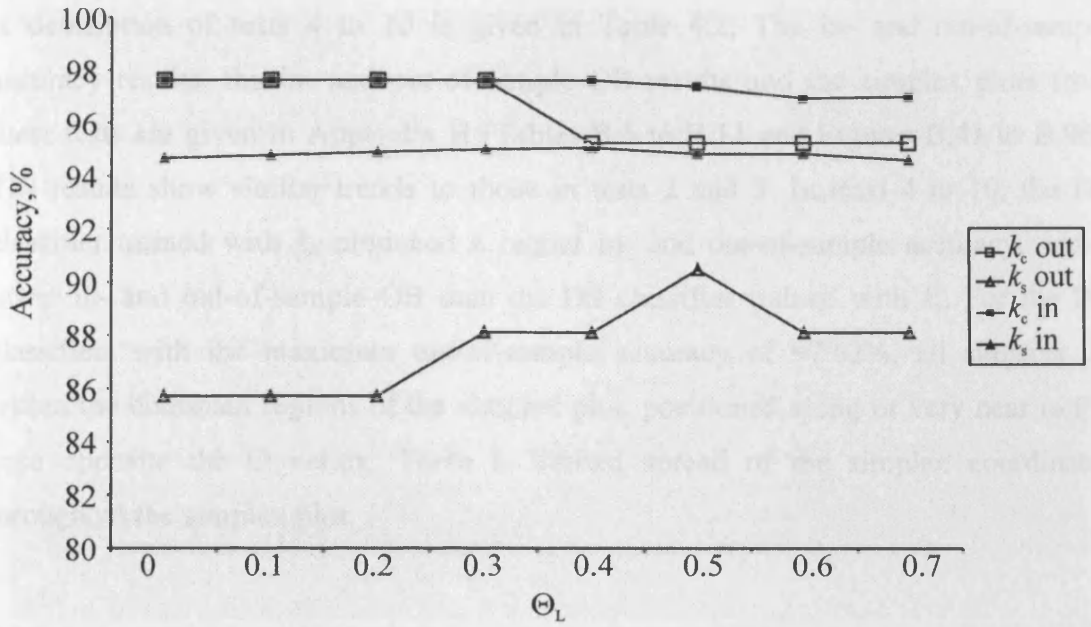


Figure 4.9 Test 3 in-sample (in) and out-of-sample (out) accuracy results showing the effect of changing Θ_L when $\Theta_U = 0.8$. The results for the two different definitions of k, k_c and k_s, are presented.

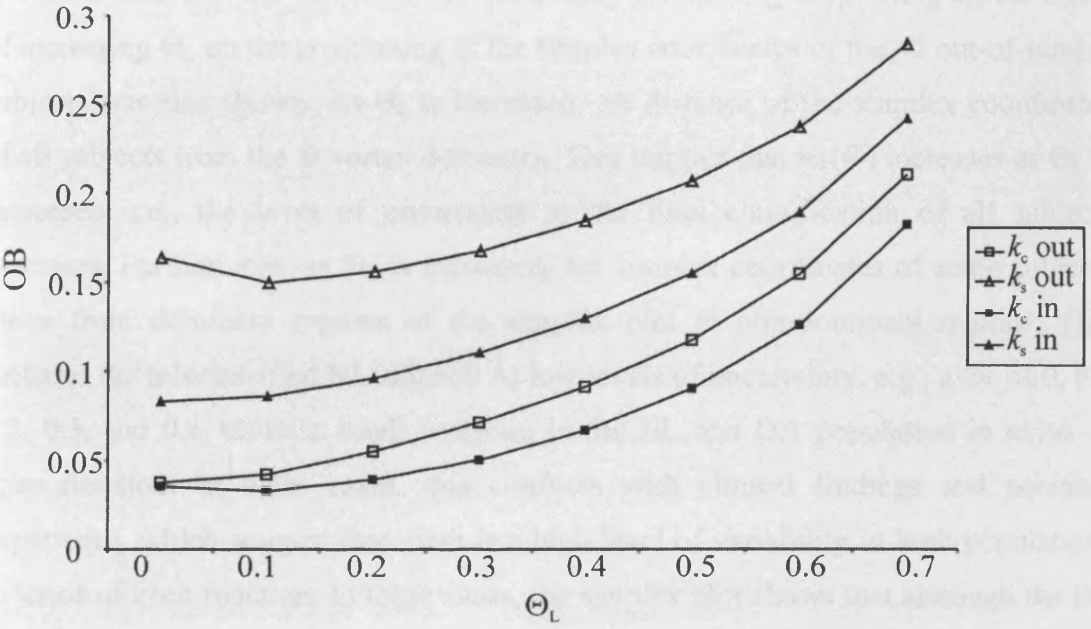


Figure 4.10 Test 3 in-sample (in) and out-of-sample (out) OB results showing the effect of changing Θ_L when $\Theta_U = 0.8$. The results for the two different definitions of k, k_c and k_s, are presented.

A description of tests 4 to 10 is given in Table 4.2. The in- and out-of-sample accuracy results, the in- and out-of-sample OB results and the simplex plots from these tests are given in Appendix B (Tables B.5 to B.11 and Figures B.41 to B.96). The results show similar trends to those in tests 2 and 3. In tests 4 to 10, the DS classifier trained with k_c produced a higher in- and out-of-sample accuracy, and a lower in- and out-of-sample OB than the DS classifier trained with k_s. For the DS classifiers with the maximum out-of-sample accuracy of 97.62%, all subjects lie within the dominant regions of the simplex plot, positioned along or very near to the base opposite the Θ vertex. There is limited spread of the simplex coordinates throughout the simplex plot.

4.1.3 Summary of results from non-optimisation method

The results from tests 1 to 10 (section 4.1.2) show that the DS classifier that uses k_c produces a higher in- and out-of-sample accuracy, and a lower in- and out-of-sample OB than the DS classifier trained with k_s. The results also show that the DS classifier is very sensitive to the choice of the uncertainty limits, $[\Theta_L, \Theta_U]$. Using k_c, the effect of increasing Θ_L on the positioning of the simplex coordinates of the 42 out-of-sample subjects was also shown. As Θ_L is increased, the distance of the simplex coordinates of all subjects from the Θ vertex decreases. This implies that $m_c(\Theta)$ increases as Θ_L is increased, i.e., the level of uncertainty in the final classification of all subjects increases. Furthermore, as Θ_L is increased, the simplex coordinates of some subjects move from dominant regions of the simplex plot to non-dominant regions. This includes the misclassified NL subject. At low levels of uncertainty, e.g., a Θ_L of 0, 0.1, 0.2, 0.3, and 0.4, there is small variation in the NL and OA population in terms of knee function. In these cases, this conflicts with clinical findings and personal experience, which suggest that there is a high level of variability in both populations in terms of knee function. In these cases, the simplex plot shows that although the DS classifier is highly accurate, it is not of practical and clinical use. At high levels of uncertainty $[\Theta_L, \Theta_U] = [0.9, 1]$, the simplex coordinates of the majority of subjects lie within the non-dominant regions. Despite having a high level of classification accuracy, the lack of dominant OA and NL classifications signify that this classifier too (the classifier with $[\Theta_L, \Theta_U] = [0.9, 1]$) has limited clinical use.

The simplex plot for the most practical DS classifier (test 1Ii) is given in Figure 4.5h. This classifier was able to classify both in- and out-of-sample subjects with an average accuracy of 97.6190%. Figure 4.5h shows that the simplex coordinates of the out-of-sample subjects are spread out within the dominant regions. This suggests a greater variability of NL subjects within the NL sample and of the OA subjects within the OA sample. The simplex coordinates of the majority of subjects lie within the dominant regions. The simplex coordinate of the misclassified NL subject lies within the non-dominant OA region of the simplex plot.

4.2 OPTIMISATION METHOD

With the optimisation method a simulated annealing (SA) algorithm is used to assign values to the DS control parameters as described in section 3.5.2. The convergence of the SA algorithm is dependent on a set of SA control parameters:

1. Starting temperature (t_s)
2. Final temperature (t_f)
3. Number of iterations per temperature (i_t)
4. Temperature reduction factor (r_t).

The significance of these SA control parameters is discussed in section 3.5.2.2. Since the way in which the SA algorithm assigns values to the DS control parameters is inherently random, the values of the DS control parameters may not always converge to the same solution. Consequently, a series of tests was conducted to investigate the influence of the SA control parameters on the values given to the DS parameters. A further series of tests was then performed to examine the effect of the SA control parameters on the performance of the DS classifier.

4.2.1 The SA control parameters

Following the work of Nolle *et al.* (2002) equation (4.4) is used to determine the starting temperature t_s, as described in section 3.5.2.2.

$$t_s = 25 \times \Delta f \tag{4.4}$$

where Δf is the difference between the current OB and the new OB.

A series of tests were conducted to determine the maximum Δf so that t_s could be established. The SA algorithm was run at three different values of t_s. The temperatures of 5, 50000 and 50000000 are values that are recorded in the literature (Sexton *et al.*, 1999). Each value of the in-sample OB calculated during the run of the SA algorithm was recorded. Δf was calculated for each step and the maximum absolute value noted. Five runs of the SA algorithm were recorded at each t_s. The results of these tests are given in Table 4.3. It can be seen that the maximum change in OB is equal to 0.0166 $\approx$ 0.02 and using equation (4.4) t_s is calculated to be 0.5.

Table 4.3 Results of tests to determine the maximum Δf

Test	Starting Temperature	Run	Max Δf
11	5	1	0.0074
		2	0.0041
		3	0.0051
		4	0.0080
		5	0.0047
12	50000	1	0.0126
		2	0.0091
		3	0.0123
		4	0.0122
		5	0.0110
13	50000000	1	0.0134
		2	0.0153
		3	0.0134
		4	0.0153
		5	0.0166

No set method was found in the literature for determining appropriate values for the remaining three SA control parameters. As such, a range of values, based around those found in the literature (Sexton *et al.*, 1999; Nolle *et al.*, 2000), was assigned to these parameters as shown in Table 4.4.

Table 4.4 Range of values used for the SA control parameters

SA control parameter	Values of parameter
Final Temperature, t_f	1×10^{-5}, 1×10^{-10}, 1×10^{-20}
Iterations Per Temperature, i_t	20, 50, 100
Temperature Reduction Factor, r_t	0.5, 0.8, 0.9

4.2.2 The effect of changing the SA parameters on the values assigned to the DS control parameters

A series of tests was conducted to establish the influence of three other SA control parameters on the values of the DS control parameters. A description of the tests is given in Table 4.5. Ten runs were performed for each part of the test. The values of k, θ, A and B were recorded for each variable. An average of these values from the 10 runs is given in Appendix B (Tables B.12 to B.23).

Following Beynon *et al.*, (2002) initial values were given to θ using the population mean, and for k using the Pearson correlation coefficient (see section 4.1.1). Initial values of 0.5 and 0.8 were assigned to A and B respectively. During optimisation, the value of B was restricted to the interval [0.8, 1], thus limiting the maximum value that can be assigned to $m(\{OA\})$ or $m(\{NL\})$ to 0.2. No restrictions were placed on the values of k, θ and A.

The effect of changing the SA control parameters on the DS control parameters is presented in sections 4.2.2.1 to 4.2.2.3. Because of the wealth of results that were recorded in tests 14 to 16, the following sections concentrate on the consequence of changing the SA parameters on the DS control parameters of four variables (v_2, v_7, v_8 and v_9) only.

Table 4.5 Description of tests carried out to determine the effect of changing the SA control parameters on the values assigned to the DS control parameters.

Test	SA control parameter changed	Part	SA control parameter value
14	t_f	a	1×10^{-5}
		b	1×10^{-10}
		c	1×10^{-20}
15	i_t	a	20
		b	50
		c	100
16	r_t	a	0.5
		b	0.8
		c	0.9

These four variables were chosen because of the work conducted in Chapter 7 (section 7.5). For the DS classifier, the ranking of variables study identified v_2 and v_8 as the two most important variables, and v_7 and v_9 as the least important variables. In Chapter 2 (section 2.6), v_2 is defined as the cadence. v_7 is defined as the VFPC1 Score, which is related to the vertical ground reaction force during 28% to 42% stance phase (mid stance) and 49% to 81% stance phase (terminal stance), namely a portion of mid-stance and the portion from heel-rise to opposite initial contact. v_8 is defined as the VFPC2 Score, which is related to the vertical ground reaction force from 2 to 27% stance phase, which is the period from loading response to mid-stance. v_9 is defined as the VFPC3 Score, which is related to the vertical ground reaction force during pre-swing. For completeness, the results for the remaining 14 variables are given in Appendix B.

4.2.2.1 Test 14 – Final temperature, t_f

The results of tests 14a, 14b and 14c to establish the effect of t_f on the values assigned to the DS control parameters are recorded in Appendix B. The average values of k, θ,

A and B for all variables are given in Tables B.12 to B.15 respectively and summarised for v_2, v_7, v_8 and v_9 in Table 4.6.

Table 4.6 The effect of changing the number of iterations per temperature, i_t on the values assigned to the DS control parameters, k, θ, A and B, for v_2, v_7, v_8 and v_9.

Variable	DS control parameter	Number of iterations per temperature, i_t		
		Test 15a	Test 15b	Test 15c
		20	50	100
v_2	k	-0.7879	-0.9018	-1.1127
	θ	49.5857	49.1393	49.4028
	A	0.4181	0.3631	0.2788
	B	0.1987	0.1986	0.1984
v_7	k	-0.0240	-0.0235	-0.0192
	θ	-4.8E-09	-4.7E-09	-3.9E-09
	A	0.4129	0.3794	0.2061
	B	0.1926	0.1964	0.1993
v_8	k	0.9526	0.9970	1.1713
	θ	-9.8E-09	-1.0E-08	-1.2E-08
	A	0.4458	0.3898	0.2851
	B	0.1967	0.1984	0.1994
v_9	k	-0.0045	-0.0046	-0.0045
	θ	-9.4E-09	-9.5E-09	-8.8E-09
	A	0.4281	0.3594	0.2498
	B	0.1911	0.1973	0.1986

The consequence of the collective changes in the values assigned to k, θ, A and B for v_2 (cadence) on the confidence values and belief values (i.e. the BOE) is given in Figure 4.11.

The mean value of k is negative for all values of t_f. The negative value of k implies that small values of v_2 offer more support to {OA} ($cf(v) \rightarrow 1$) whilst large values offer more support to {NL} ($cf(v) \rightarrow 0$). As t_f is decreased, the absolute mean value of k increases. Such an increase in the absolute mean value of k has the effect of decreasing the range of v_2 for which the confidence function, $cf(v)$ is near 0.5. This implies that more of the v_2 measurements will transform to the extreme $cf(v)$ values of zero or one. Consequently, more $cf(v)$ values will be transformed into a BOE in which maximum values are assigned to either $m(\{OA\})$ or $m(\{NL\})$.

The mean value of θ decreases as t_f is decreased from 1×10^{-5} to 1×10^{-10} and then increases as t_f is decreased further to 1×10^{-20}. θ determines the value of v_2 for which $cf(v) = 0.5$. The effect of these changes in θ on the confidence values is minimal.

The mean value of A decreases as t_f is decreased. The change in the value of A has a significant effect on the BOE values. The decrease in the value of A results in a decrease in the maximum value assigned to $m(\Theta)$ and a decrease in the variation of $m(\Theta)$ with $cf(v)$. The change in the value of A has no effect on the extreme $cf(v)$ values of zero and one. However, the change is more significant for the cases where the values of $cf(v)$ are near to 0.5. In these cases, the values assigned to $m(\{OA\})$ and $m(\{NL\})$ are increased and the value of $m(\Theta)$ decreased. This implies that values of v_2 which are transformed to $cf(v)$ of near 0.5 offer more certainty to the final classification of subjects.

The mean value of B decreases slightly as t_f is decreased. However, the small change in B has minimal effect on the BOE values, because for each final temperature, t_f the maximum value assigned to $m(\{OA\})$ and $m(\{NL\}) \approx 0.2$.

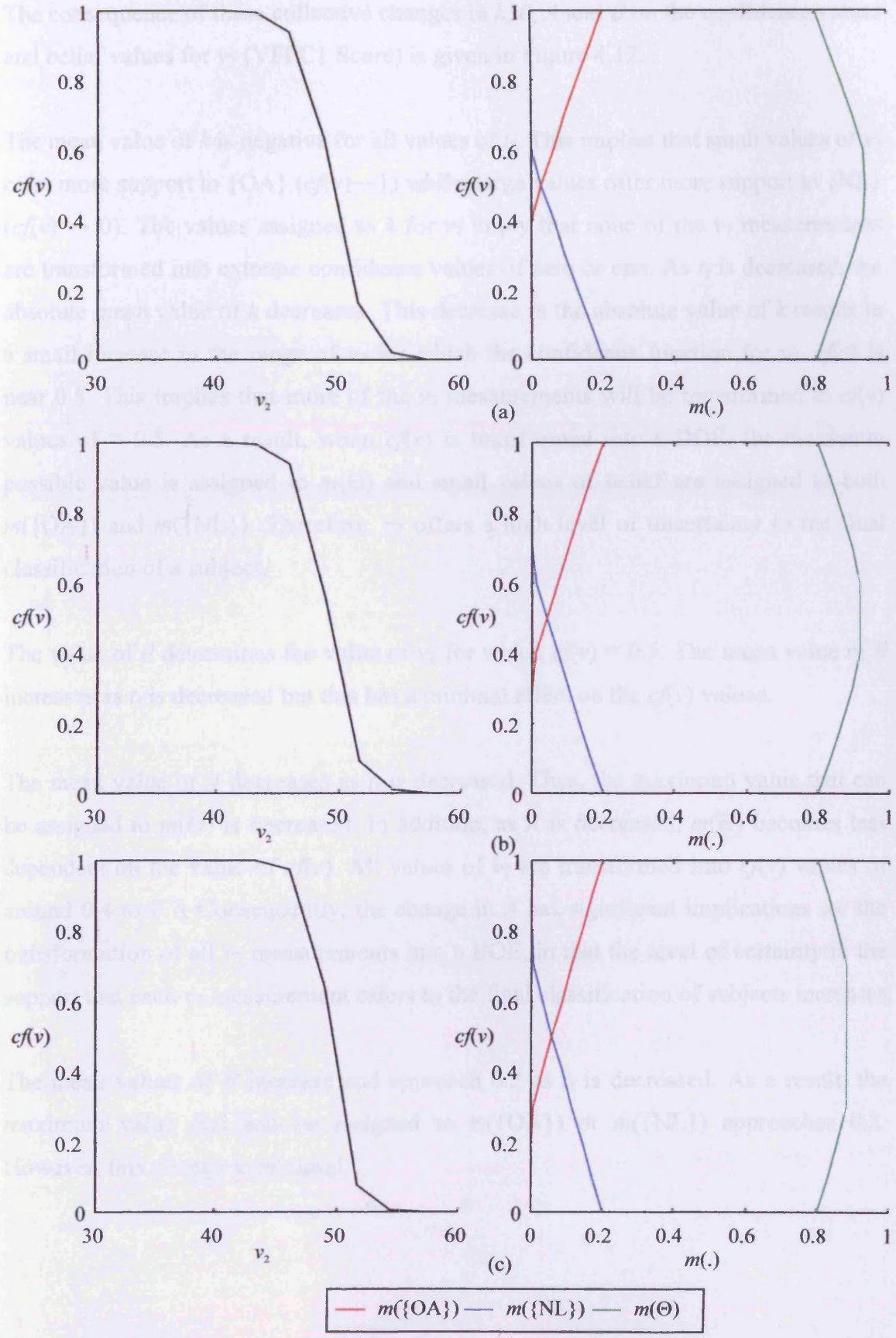


Figure 4.11 The effect of decreasing the final temperature, t_f on the confidence factor and BOE for v_2. (a) $t_f = 1\times10^{-5}$; (b) $t_f = 1\times10^{-10}$; (c) $t_f = 1\times10^{-20}$.

The consequence of these collective changes in k, θ, A and B on the confidence values and belief values for v_7 (VFPC1 Score) is given in Figure 4.12.

The mean value of k is negative for all values of t_f. This implies that small values of v_7 offer more support to {OA} ($cf(v) \rightarrow 1$) whilst large values offer more support to {NL} ($cf(v) \rightarrow 0$). The values assigned to k for v_7 imply that none of the v_7 measurements are transformed into extreme confidence values of zero or one. As t_f is decreased, the absolute mean value of k decreases. This decrease in the absolute value of k results in a small increase in the range of v_7 for which the confidence function for v_7, $cf(v)$ is near 0.5. This implies that more of the v_7 measurements will be transformed to $cf(v)$ values of ≈ 0.5. As a result, when $cf(v)$ is transformed into a BOE, the maximum possible value is assigned to $m(\Theta)$ and small values of belief are assigned to both $m(\{OA\})$ and $m(\{NL\})$. Therefore, v_7 offers a high level of uncertainty to the final classification of a subject.

The value of θ determines the value of v_2 for which $cf(v) = 0.5$. The mean value of θ increases as t_f is decreased but this has a minimal effect on the $cf(v)$ values.

The mean value of A decreases as t_f is decreased. Thus, the maximum value that can be assigned to $m(\Theta)$ is decreased. In addition, as A is decreased, $m(\Theta)$ becomes less dependent on the value of $cf(v)$. All values of v_7 are transformed into $cf(v)$ values of around 0.4 to 0.7. Consequently, the change in A has significant implications for the transformation of all v_7 measurements into a BOE, in that the level of certainty in the support that each v_7 measurement offers to the final classification of subjects increases.

The mean values of B increase and approach 0.2 as t_f is decreased. As a result, the maximum value that can be assigned to $m(\{OA\})$ or $m(\{NL\})$ approaches 0.2. However, this change is minimal.

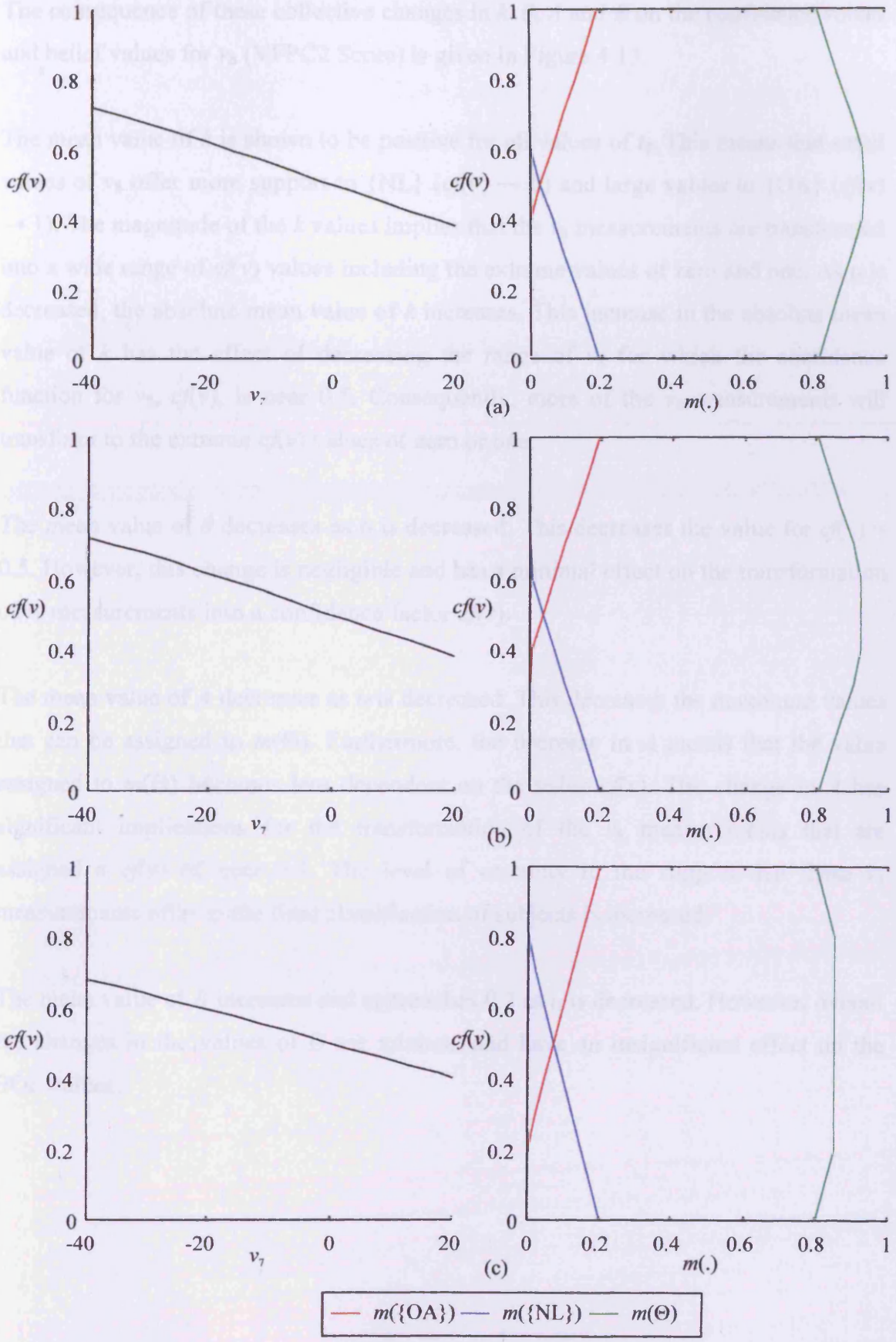


Figure 4.12 The effect of decreasing the final temperature, t_f on the confidence factor and BOE for v_7. (a) $t_f = 1\times10^{-5}$; (b) $t_f = 1\times10^{-10}$; (c) $t_f = 1\times10^{-20}$.

The consequence of these collective changes in k, θ, A and B on the confidence values and belief values for v_8 (VFPC2 Score) is given in Figure 4.13.

The mean value of k is shown to be positive for all values of t_f. This means that small values of v_8 offer more support to {NL} ($cf(v) \rightarrow 0$) and large values to {OA} ($cf(v) \rightarrow 1$). The magnitude of the k values implies that the v_8 measurements are transformed into a wide range of $cf(v)$ values including the extreme values of zero and one. As t_f is decreased, the absolute mean value of k increases. This increase in the absolute mean value of k has the effect of decreasing the range of v_8 for which the confidence function for v_8, $cf(v)$, is near 0.5. Consequently, more of the v_8 measurements will transform to the extreme $cf(v)$ values of zero or one.

The mean value of θ decreases as t_f is decreased. This decreases the value for $cf(v) = 0.5$. However, this change is negligible and has a minimal effect on the transformation of v_8 measurements into a confidence factor $cf(v)$.

The mean value of A decreases as t_f is decreased. This decreases the maximum values that can be assigned to $m(\Theta)$. Furthermore, the decrease in A means that the value assigned to $m(\Theta)$ becomes less dependent on the value $cf(v)$. The change in A has significant implications for the transformation of the v_8 measurements that are assigned a $cf(v)$ of near 0.5. The level of certainty in the support that these v_8 measurements offer to the final classification of subjects is increased.

The mean value of B increases and approaches 0.2 as t_f is decreased. However, overall the changes in the values of B are minimal and have an insignificant effect on the BOE values.

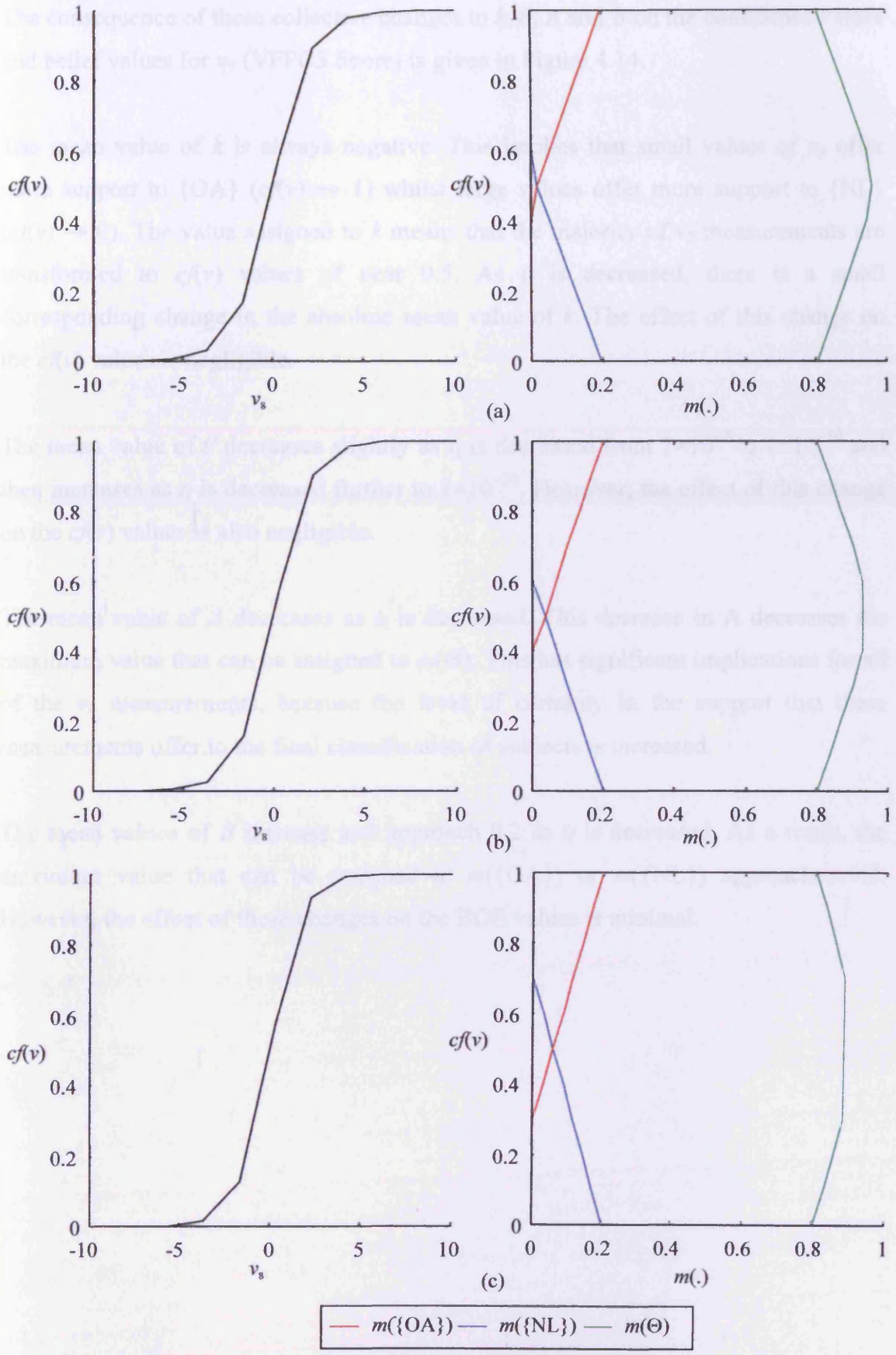


Figure 4.13 The effect of decreasing the final temperature, t_f on the confidence factor and BOE for v_8. (a) $t_f = 1\times10^{-5}$; (b) $t_f = 1\times10^{-10}$; (c) $t_f = 1\times10^{-20}$.

The consequence of these collective changes to k, θ, A and B on the confidence values and belief values for v_9 (VFPC3 Score) is given in Figure 4.14.

The mean value of k is always negative. This implies that small values of v_9 offer more support to {OA} ($cf(v) \rightarrow 1$) whilst large values offer more support to {NL} ($cf(v) \rightarrow 0$). The value assigned to k means that the majority of v_9 measurements are transformed to $cf(v)$ values of near 0.5. As t_f is decreased, there is a small corresponding change in the absolute mean value of k. The effect of this change on the $cf(v)$ values is negligible.

The mean value of θ decreases slightly as t_f is decreased from 1×10^{-5} to 1×10^{-10} and then increases as t_f is decreased further to 1×10^{-20}. However, the effect of this change on the $cf(v)$ values is also negligible.

The mean value of A decreases as t_f is decreased. This decrease in A decreases the maximum value that can be assigned to $m(\Theta)$. This has significant implications for all of the v_9 measurements, because the level of certainty in the support that these measurements offer to the final classification of subjects is increased.

The mean values of B increase and approach 0.2 as t_f is decreased. As a result, the maximum value that can be assigned to $m(\{OA\})$ or $m(\{NL\})$ approaches 0.2. However, the effect of these changes on the BOE values is minimal.

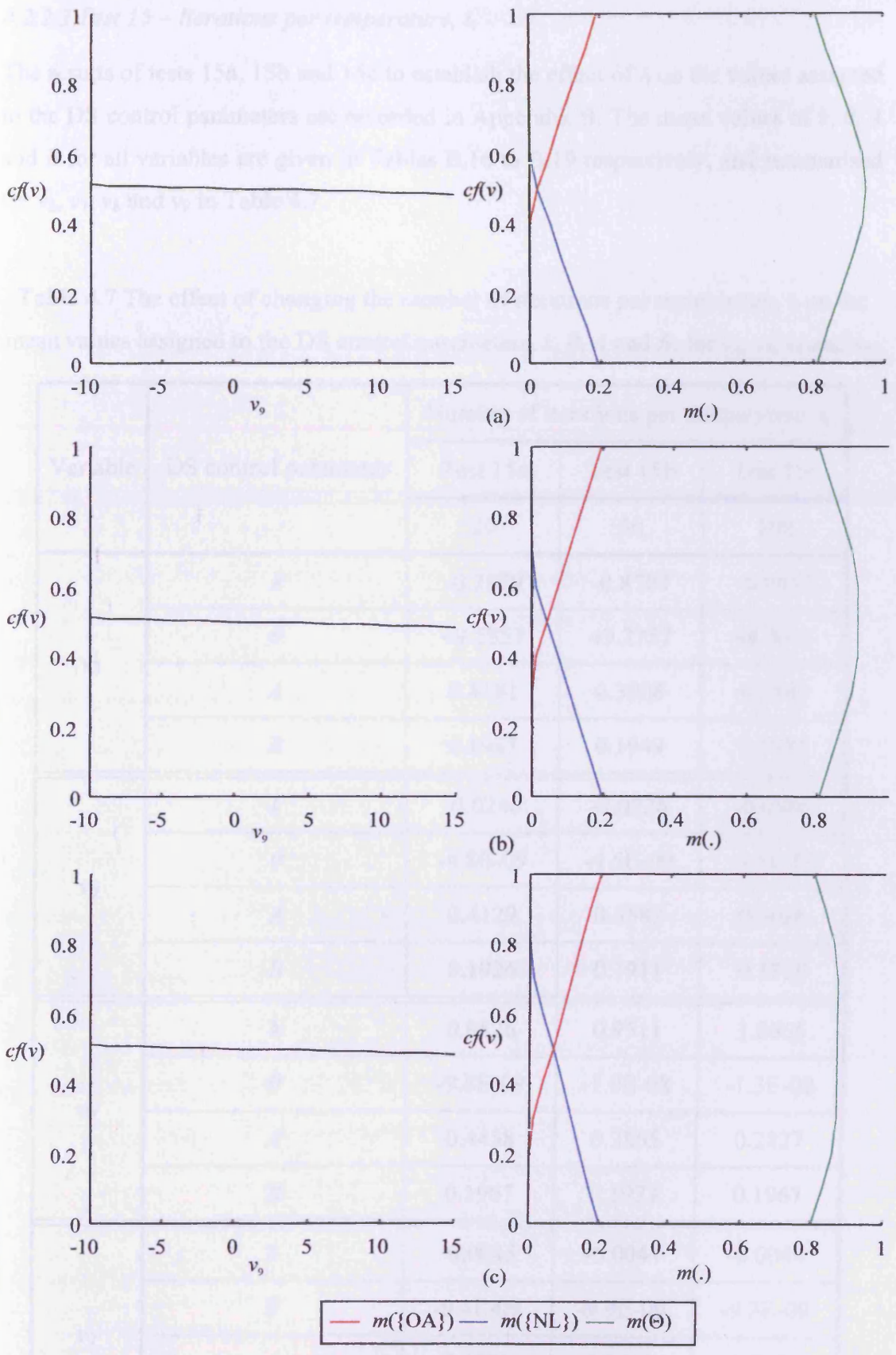


Figure 4.14 The effect of decreasing the final temperature, t_f on the confidence factor and BOE for v_9. (a) $t_f = 1\times10^{-5}$; (b) $t_f = 1\times10^{-10}$; (c) $t_f = 1\times10^{-20}$.

4.2.2.2 Test 15 – Iterations per temperature, i_t

The results of tests 15a, 15b and 15c to establish the effect of i_t on the values assigned to the DS control parameters are recorded in Appendix B. The mean values of k, θ, A and B for all variables are given in Tables B.16 to B.19 respectively, and summarised for v_2, v_7, v_8 and v_9 in Table 4.7.

Table 4.7 The effect of changing the number of iterations per temperature, i_t on the mean values assigned to the DS control parameters, k, θ, A and B, for v_2, v_7, v_8 and v_9.

Variable	DS control parameter	Number of iterations per temperature, i_t		
		Test 15a	Test 15b	Test 15c
		20	50	100
v_2	k	-0.7879	-0.8704	-0.9912
	θ	49.5857	49.2757	48.3011
	A	0.4181	0.3908	0.2840
	B	0.1987	0.1949	0.1971
v_7	k	-0.0240	-0.0226	-0.0180
	θ	-4.8E-09	-4.5E-09	-4.5E-09
	A	0.4129	0.3587	0.3086
	B	0.1926	0.1911	0.1910
v_8	k	0.9526	0.9511	1.0665
	θ	-9.8E-09	-1.0E-08	-1.3E-08
	A	0.4458	0.3855	0.2827
	B	0.1967	0.1973	0.1967
v_9	k	-0.0045	-0.0044	-0.0049
	θ	-9.4E-09	-9.9E-09	-9.3E-09
	A	0.4281	0.3609	0.2499
	B	0.1911	0.1905	0.1859

For v_2 (cadence), the mean value of k is negative for all values of i_t. As i_t is increased, the absolute mean value of k increases. The mean value of θ decreases as i_t is increased. The mean value of A decreases as i_t is increased. As i_t is increased from 20 to 50 and subsequently from 50 to 100, the mean value of B decreases and then increases.

For v_7 (VFPC1 Score), the mean value of k is negative for all values of i_t. As i_t is increased, the absolute mean value of k decreases. The mean value of θ increases as i_t is increased from 20 to 50 and then remains constant as i_t is increased further to 100. The mean value of A decreases as i_t is increased. The mean value of B decreases as i_t is increased.

For v_8 (VFPC2 Score), the mean value of k is positive for all values of i_t. The absolute mean value of k for v_8 decreases slightly and subsequently increases as i_t is increased from 20 to 50 and then from 50 to 100. The mean value of θ decreases as i_t is increased. The mean value of A decreases as i_t is increased. As i_t is increased from 20 to 50 the mean value of B increases. Increasing i_t to 100 results in a decrease in the mean value of B.

For v_9 (VFPC3 Score), the mean value of k is negative for all values of i_t. The mean value of θ for v_9 decreases as it is increased from 20 to 50 and then increases as it is increased further to 100. The mean value of A decreases as i_t is increased. The mean value of B decreases as i_t is increased.

The effect of the combined changes in the DS control parameters on the confidence factor and the body of evidence for variables v_2, v_7, v_8 and v_9 are shown in Appendix B. The consequences of these changes are similar to those described in test 15.

4.2.2.3 Test 16 – Temperature reduction factor, r_t

The results of tests 16a, 16b and 16c to establish the effect of r_t on the values assigned to the DS control parameters are recorded in Appendix B. The average values of k, θ, A and B are given in Tables B.20 to B.23 respectively and are summarised for the variables v_2, v_7, v_8 and v_9 in Table 4.8.

Table 4.8 The effect of changing the number of iterations per temperature, r_t on the mean values assigned to the DS control parameters, k, θ, A and B, for v_2, v_7, v_8 and v_9.

Variable	DS control parameter	Number of iterations per temperature, i_t		
		Test 15a	Test 15b	Test 15c
		20	50	100
v_2	k	-0.7879	-0.9713	-0.9512
	θ	49.5857	49.1174	48.9006
	A	0.4181	0.3505	0.2573
	B	0.1987	0.1949	0.1988
v_7	k	-0.0240	-0.0202	-0.0177
	θ	-4.8E-09	-4.8E-09	-3.9E-09
	A	0.4129	0.3484	0.2232
	B	0.1926	0.1892	0.1932
v_8	k	0.9526	0.9687	1.2214
	θ	-9.8E-09	-1.0E-08	-1.3E-08
	A	0.4458	0.3893	0.2724
	B	0.1967	0.1929	0.1982
v_9	k	-0.0045	-0.00424	-0.00431
	θ	-9.4E-09	-8.5E-09	-8.5E-09
	A	0.4281	0.3265	0.1715
	B	0.1911	0.1935	0.1961

For v_2, the mean value of k is negative, for all values of r_t. As r_t is increased from 0.5 to 0.8 and then from 0.8 to 0.9, the absolute mean value of k increases from 0.7879 to 0.9713 and then decreases slightly to 0.9512. The mean value of θ for v_2 decreases as r_t is increased. For v_2, the mean value of A decreases as r_t is increased. The mean

values of B decrease as r_t is increased from 0.5 to 0.8 and then increase as r_t is increased from 0.8 to 0.9.

For v_7, the mean value of k is negative for all values of r_t. As r_t is increased the absolute mean value of k decreases. The mean value of θ remains constant as r_t is increased from 0.5 to 0.8 and then increases as r_t is increased to 0.9. The mean value of A decreases as r_t is increased. The mean values of B decrease as r_t is increased from 0.5 to 0.8 and then increase as r_t is increased from 0.8 to 0.9.

For v_8, the mean value of k is always positive, despite the change in r_t. As r_t is increased so also the absolute mean value of k increases. The mean of θ increases as r_t is increased. The mean value of A decreases as r_t is increased. The mean values of B decrease as r_t is increased from 0.5 to 0.8 and then increase as r_t is increased from 0.8 to 0.9.

For v_9, the mean value of k is always negative. As r_t is increased, the absolute mean value of k decreases slightly and then increases slightly as r_t is increased. The mean of θ increases as r_t is increased from 0.5 to 0.8. As r_t is increased further to 0.9, the mean value of θ remains constant. The mean value of A decreases as r_t is increased. The mean value of B increases as r_t is increased.

The effect of the combined changes in the DS control parameters on the confidence factor and the body of evidence for variables v_2, v_7, v_8 and v_9 are shown in Appendix B. The consequences of these changes are similar to those described in tests 15 and 16.

4.2.3 Effect of changing the SA parameters on the performance of the DS classifier

A series of tests was conducted to determine the effect of changing the SA parameters on the performance of the DS classifier. A description of these tests is given in Table 4.9. Each part of the test was repeated ten times. For each run the average in- and out-of sample accuracy and average in- and out-of-sample OB were calculated using a LOO approach.

Table 4.9 Description of tests carried out to determine best training strategy.

Test	SA control parameter changed
17	t_f
18	i_t
19	r_t

Part	SA control parameter value
a	1×10^{-5}
b	1×10^{-10}
c	1×10^{-20}
a	20
b	50
c	100
a	0.5
b	0.8
c	0.9

4.2.3.1 Test 17 – Final temperature, t_f

The results of tests 17a to 17c to establish the effect of changing t_f on the in- and out-of-sample accuracy and OB are recorded in Appendix B (Tables B.24 to B.26). The average and standard deviation of the ten training runs are summarised for the in- and out-of-sample accuracies in Table 4.10 and for the OB in Table 4.11.

Table 4.10 Test 17 results. Summary of the average in-sample and out-of-sample accuracy, % for three different final temperatures, t_f.

Part	t_f	In-sample accuracy, %		Out-of-sample accuracy, %	
		Mean	Standard Deviation	Mean	Standard Deviation
a	1×10^{-5}	96.86	0.22	95.24	1.11
b	1×10^{-10}	96.28	0.23	94.52	1.61
c	1×10^{-20}	95.36	0.06	93.34	1.00

Table 4.11 Test 17 results. Summary of the average in-sample and out-of-sample OB for three different final temperatures, t_f.

Part	t_f	In-sample OB		Out-of-sample OB	
		Mean	Standard Deviation	Mean	Standard Deviation
a	1×10^{-5}	0.3363	0.0001	0.3313	0.0035
b	1×10^{-10}	0.3363	0	0.3070	0.0035
c	1×10^{-20}	0.3363	0	0.2857	0.0014

As t_f is decreased, that is the SA algorithm runs for a longer period, the average in- and out-of-sample accuracy of the DS method decreases. The in-sample accuracy of the DS method is always greater than the out-of-sample accuracy. The standard deviation values indicate that the variability of both the in- and out-of-sample accuracies is seen to increase and then decrease as t_f is decreased, with substantially more variability in the out-of-sample results.

For each value of t_f, the in-sample OB converged to a value of 0.3363. However, as the t_f is decreased the value of the out-of-sample OB decreased. Decreasing t_f from 1×10^{-5} to 1×10^{-10} did not change the standard deviation of the out-of-sample OB. A further decrease in t_f to 1×10^{-20} had the result of decreasing the standard deviation.

4.2.3.2 Test 18 – Number of iterations per temperature, i_t

The results of tests 18a, 18b and 18c to establish the effect of i_t on the in- and out-of-sample accuracy and in- and out-of-sample OB are recorded in Appendix B (Tables B.27 to B.29 respectively). The average and standard deviation of the ten training runs for the in- and out-of-sample accuracy are summarised in Table 4.12 and for the in-sample and out-of-sample OB in Table 4.13.

Table 4.12 Test 18 results. Summary of average in-sample and out-of-sample accuracy, % for three different numbers of iterations per temperature, i_t.

Part	i_t	In-sample accuracy %		Out-of-sample accuracy, %	
		Mean	Standard Deviation	Mean	Standard Deviation
a	20	96.86	0.22	95.24	1.11
b	50	96.40	0.13	95.48	0.75
c	100	95.75	0.14	93.81	2.01

Table 4.13 Test 18 results. Summary of average in-sample and out-of-sample OB for three different numbers of iterations per temperature, i_t.

Part	i_t	In-sample OB		Out-of-sample OB	
		Mean	Standard Deviation	Mean	Standard Deviation
a	20	0.3363	0.0001	0.3313	0.0035
b	50	0.3363	0	0.3128	0.0024
c	100	0.3363	0	0.2942	0.0032

As i_t is increased, the average in-sample accuracy decreases. The out-of-sample accuracy increases slightly as the number of iterations is increased from 20 to 50, and then decreases as the number of iterations is increased further to 100. The in-sample accuracy of the DS method is always greater than the out-of-sample accuracy. The variability of both the in- and out-of-sample accuracies is seen to decrease and then increase as i_t is increased, which is reflected in the values of the standard deviation with the out-of-sample results showing more variability.

For each value of i_t, the in-sample OB converges to a value of 0.3363. However as i_t is decreased the value of the out-of-sample OB decreases. The standard deviation of the

out-of-sample OB decreases as i_t is increased from 20 to 50 and then increases as i_t is increased to 100.

4.2.3.3 Test 19 – Temperature reduction factor, r_t

The results of tests 19a, 19b and 19c to establish the effect of r_t on the in- and out-of-sample accuracy and in- and out-of-sample OB are recorded in Appendix B (Tables B.30 to B.32 respectively). The average and standard deviation of the ten training runs for the in-sample and out-of-sample accuracy are summarised in Table 4.14 and for the in-sample and out-of-sample OB in Table 4.15.

Table 4.14 Test 19 results. Summary of average in-sample and out-of-sample accuracy, % for three different temperature reduction factors, r_t.

Part	r_t	In-sample accuracy, %		Out-of-sample accuracy, %	
		Mean	Standard Deviation	Mean	Standard Deviation
a	0.5	96.86	0.22	95.24	1.11
b	0.8	96.27	0.20	94.52	1.15
c	0.9	95.49	0.15	94.29	1.23

Table 4.15 Test 19 results. Summary of average in-sample and out-of-sample OB for three different temperature reduction factors, r_t.

Part	r_t	In-sample OB		Out-of-sample OB	
		Mean	Standard Deviation	Mean	Standard Deviation
a	0.5	0.3363	0.0001	0.3313	0.0035
b	0.8	0.3363	0	0.3082	0.0051
c	0.9	0.3363	0	0.2871	0.0019

As the r_t is increased both the in- and out-of-sample accuracy decrease. The in-sample accuracy of the DS method is always greater than the out-of-sample accuracy. The variability in the in-sample accuracy is seen to decrease as r_t is increased, which is reflected in the values of the standard deviation. In contrast, the standard deviation of the out-of-sample accuracy increases as r_t is increased. Comparing the standard deviation of the in-sample accuracy with the out-of-sample accuracy shows that the out-of-sample results are more variable.

For each value of r_t, the in-sample OB converges to a value of 0.3363. However as i_t is decreased the value of the out-of-sample OB decreases. The standard deviation of the out-of-sample OB decreases as i_t is increased from 20 to 50 and then increases as i_t is increased to 100. The standard deviation of the out-of-sample OB increases as r_t is increased from 0.5 to 0.8 and then decreases as r_t is increased further to 0.9.

4.2.4 Summary of results using the optimisation method

Tests 14, 15 and 16 showed the effect of changing the SA parameters on the values assigned to the DS control parameters, k, θ, A and B. In limiting the maximum value that could be assigned to $m(\{\text{OA}\})$ or $m(\{\text{NL}\})$ to 0.2, changing the SA parameters had the greatest effect on the value assigned to A. Running the SA algorithm for a longer period always resulted in a decrease in the value of A. Consequently, the maximum value that could be assigned to $m(\Theta)$ was decreased. This has a significant effect on the measurements that are transformed into $cf(v)$ values of near 0.5, because the level of certainty in the support that these measurements offer to the final classification of subjects is increased.

The results of tests 17, 18 and 19 show that the SA algorithm always converged to an in-sample OB value of 0.3363 regardless of the SA control parameters used. Running the SA algorithm for a longer period – i.e. reducing the final temperature, increasing the number of iterations per temperature or increasing the temperature reduction factor – reduced the value of the out-of-sample OB. This implies that the level of uncertainty associated with the final classification of subjects decreases. However, running the SA algorithm for a longer period resulted in a decrease in the in- and out-

of-sample accuracy. The results of tests 14 to 16 suggest that this is a direct consequence of the decrease in the values assigned to A.

Together, the results of tests 14 to 19 suggest that in the case where the limits of A = [0, 0.5] and B = [0, 0.2], the SA algorithm should be run for a short period. The results also suggest that the lower limit placed on A should be increased to restrict the level of belief that is assigned to measurements that are transformed into $cf(v)$ values of 0.5.

A summary of the best DS classifiers for the non-optimisation method (tests 1 to 10) and the optimisation method (tests 17 to 19) is given in Table 4.16. For the boundary conditions used in this study, the non-optimisation method produced a superior classifier to the optimisation method in terms of in- and out-of-sample accuracy. However, using different boundary conditions on A and B may result in an improved classifier when using the optimisation technique. Such an investigation was beyond the scope of this thesis but should be investigated in further studies.

In this study, assigning values to the DS control parameters using the non-optimisation method produces a superior classifier in terms of in-sample and out-of-sample accuracy, than using the optimisation method. Consequently, the non-optimisation DS classifier was used in the following studies in Chapters 5 to 7.

Table 4.16 A comparison of the best classifiers from the non-optimisation and optimisation methods in terms of in-sample and out-of-sample accuracy, %.

Method	Accuracy, %	
	In-sample	Out-of-sample
Non-optimisation	97.62	97.62
Optimisation	96.86	95.24

In this chapter, a necessary study has been undertaken to investigate the validity of the DS classifier. Extensive results have been presented and have provided a valuable insight into the sensitivity of the method to the various control parameters and their

effect on the performance of the DS classifier. Conclusions based on this chapter and suggestions for further work are given in Chapter 8.

CHAPTER 5

FEATURE SELECTION

This chapter investigates the feasibility of using automatic feature selection in conjunction with the Dempster-Shafer (DS) classifier. The chapter begins in sections 5.1 and 5.2 with an introduction to the subject of feature selection. Subsequently, three feature selection methods are described in detail and their suitability for use with the DS classifier is investigated in sections 5.3 to 5.5. The three chosen methods, Stepwise Linear Discriminant Analysis (SLDA), Sequential Selection Methods (SSM) and Genetic Algorithms (GA), have been used in previous gait classification studies. In section 5.6 a comparison is made between the three DS classifiers using feature selection with the DS classifier that does not use feature selection (see Chapter 4, section 4.1.2.1, test 1Ii). This comparison is made in order to address the question of whether it is best to use an automated approach or to rely on expert clinical opinion when choosing the input variables that should be used in the classification of osteoarthritic (OA) and normal (NL) knee function.

5.1 INTRODUCTION

It is widely appreciated within the fields of artificial intelligence and pattern recognition that the appropriate selection of input variables (or features; these terms are used interchangeably) is an integral part of classifier design (Duda *et al.*, 2001; Siedlecki and Sklansky, 1988; Raymer *et al.*, 2000). This selection process, commonly known as feature selection, can have a significant influence on the performance of a classifier in terms of both cost and accuracy (Raymer *et al.*, 2000). Reducing the number of inputs to a classifier inherently reduces the computational time needed to train and evaluate it (Dash and Liu, 1997) and limits the number of features that need to be measured at the outset (Raymer *et al.*, 2000; Jain and Zongker, 1997). Furthermore, for a given amount of training examples, reducing the number of features can improve the estimate of the classifier's performance (Siedlecki and Sklansky, 1988; Raymer *et al.*, 2000). The use of feature selection can enhance the classifier's performance (Su and Wu, 2000). Finally, automatic feature selection is

a very useful tool given that it is not always obvious *a priori,* as to which variables are most useful in determining an object's class (Raymer *et al.*, 2000).

Given a set of p input variables, $V = \{v_1, v_2, ..., v_p\}$, it would be ideal to conduct an exhaustive search through the entire set of subsets of V, to establish the best subset according to some evaluation function. However, in practice, this is computationally infeasible since there are 2^p possible subsets of V. For example, for $p = 18$ there are $2^{18} \approx 3 \times 10^5$ subsets of V. Consequently, sub-optimal feature selection methods have become commonplace in classifier design

Despite the rewards of feature selection and its success in other fields, only a handful of gait classification studies have taken advantage of it (Chao *et al.*, 1980; Lafuente *et al.*, 1998; Astephen *et al.*, 2002a; Astephen *et al.*, 2002b; Su and Wu, 2000). These studies are now discussed in brief.

Chao *et al.* (1980) employed a Stepwise Regression method (Kendall and Stuart, 1973) to identify a subset of gait variables that were most significant in providing discriminative power in separating post-operative total-knee replacement and NL subjects. Using this feature selection technique the number of features to be used in their analysis was reduced from forty-three to eight.

Subsequent to performing Principal Component Analysis on a combined data set of constant gait parameters and time-varying gait waveforms, Astephen *et al.* (2002a, 2002b) employed a SLDA to identify a subset of Principal Components most able to discriminate between OA and NL subjects. The discriminatory power of these Principal Components was then established through examination of the weighting coefficients of the linear discriminant function.

Prior to designing an Artificial Neural Network (ANN) classifier, Lafuente *et al.* (1998) utilised SLDA to reduce the size of their input features vector. The features were ranked according to the order in which they were entered into the stepwise analysis. The most discriminatory features were then used as inputs to the classifier. This selection process reduced the number of input features from forty-one to thirty.

Su and Wu (2000) used a GA to eliminate features that did not add discriminatory power to their ANN classifier. Using the GA improved the performance of the ANN in classifying NL and ankle arthrodesis subjects. Using feature selection increased the classification accuracy from 89.7 to 98.7%.

This introduction has highlighted the value of automatic feature selection when designing a classification method. The benefits of using feature selection in conjunction with the DS classifier will now be investigated. Sections 5.3 to 5.5 give a description of three different feature selection methods that were chosen for use with the DS classifier:

1. Stepwise Linear Discriminant Analysis (SLDA) (Lafuente *et al.*, 1998; Astephen *et al.*, 2002a, 2002b) (see section 5.3)
2. Sequential Selection Methods (SSM) (see section 5.4)
3. Genetic Algorithms (GA) (Su and Wu, 2000) (see section 5.5)

Firstly, the concept of an objective function is introduced in section 5.2.

5.2 OBJECTIVE FUNCTION

An evaluation or objective function is utilised to determine the performance of the feature selection algorithm. Dash and Liu (1997) give a description of the different types of evaluation function used in classification problems. Briefly, these evaluation functions are grouped into five categories, namely, distance measures, information measures, dependence measures, consistency measures and classifier error rate measures. The type of evaluation function used will depend on the goal of feature selection, which is often application specific.

For the present study, the aim of feature selection is to select a subset of input variables (from the entire set of input variables listed in Table 5.1) that primarily improves the classification accuracy of the classifier and secondarily reduces the amount of uncertainty associated with the final classification of a subject (Jones *et al.* 2003b). In Chapter 3, the classification accuracy is calculated from the leave-one-out classifier error rate and the level of uncertainty using a Euclidean distance measure,

defined OB. Thus, in this study, it follows that a suitable evaluation function for feature selection, OB_{FS} should consist of a combination of the classifier error rate and the OB distance measure. As such, OB_{FS} is defined as:

$$OB_{FS} = \alpha + OB_{IN} \tag{5.1}$$

where α is defined as the average in-sample classification error (equal to 100 minus the average in-sample classification accuracy) and OB_{IN} is the average in-sample OB. The aim of feature selection is to minimise this function. Thus, a DS classifier with a high classification accuracy will have a lower associated OB_{FS} than a DS classifier with a low classification accuracy. Furthermore, for two DS classifiers with the same classification accuracies but with different levels of uncertainty, the OB_{FS} will be less for the classifier with a lower level of associated uncertainty.

Table 5.1 Variables used in the analysis.

Variable, v_i	Variable Description
v_1	BMI
v_2	Cadence
v_3	Stance
v_4	APFPC1 Score
v_5	APFPC2 Score
v_6	APFPC3 Score
v_7	VFPC1 Score
v_8	VFPC2 Score
v_9	VFPC3 Score
v_{10}	FERPC1 Score
v_{11}	FERPC2 Score
v_{12}	AARPC1 Score
v_{13}	AARPC2 Score
v_{14}	AARPC3 Score
v_{15}	IERPC1 Score
v_{16}	ML Width
v_{17}	AP Width
v_{18}	Thigh Girth

5.3 STEPWISE LINEAR DISCRIMINANT ANALYSIS (SLDA)

5.3.1 SLDA Theory

Using Linear Discriminant Analysis (LDA) a set of variables are weighted and linearly combined in such a way that forces two or more groups to be as distinct as possible. In application to the current study in classifying NL and OA knee function, let us consider the two-group classification problem. For this case, LDA yields a single discriminant function of the form

$$D = d_0 + d_1 v_1 + d_2 v_2 + \ldots + d_p v_p \tag{5.2}$$

where D is the discriminant score, p the number of variables, d_0 a constant factor, d_i (i = 1: p) the discriminant function coefficient and v_i (i = 1: p) the input variable. For an explanation of how the discriminant function coefficients are calculated, the reader is directed to Green (1978). Using SLDA this discriminant function is built iteratively. At each stage, the variable that maximises the difference between the two groups (OA and NL) is added. This difference is measured in terms of the Wilks' lambda statistic. For the first stage when only one variable is entered into the analysis, Wilks' lambda, Λ, is defined as the ratio

$$\Lambda = \frac{|SS_w|}{|SS_t|} \tag{5.3}$$

where SS_w is the within-group sum of squares matrix and SS_t the total sum of squares matrix. $SS_w = SS_{OA} + SS_{NL}$, where SS_{OA} and SS_{NL} are the total sum of squares matrices for the OA and NL groups respectively. For the second stage onwards, when there is more than one variable in the analysis, lambda is defined as:

$$\Lambda = \frac{|SSCP_w|}{|SSCP_t|} \tag{5.4}$$

where $SSCP_w$ is the within-group sums of squares and cross products matrix and $SSCP_t$ is the total sums of squares and cross products matrix (Green, 1978). Lambda has a value in the range [0, 1]; a value close to zero indicates that the OA and NL group means are different whilst values close to one indicate that the group means are not different. Thus, at each step the variable that minimises the overall Wilks' lambda is entered. Using this stepwise procedure, the input variables can be ranked in order of importance by noting the sequence in which they were included in the SLDA (Lafuente *et al.*, 1998), with the variable entered first considered more important than that entered second etc.

5.3.2 SLDA Results

An SLDA procedure was implemented in SPSS 11 (SPSS Inc.). At each step, the variable that minimised the overall Wilks' lambda was entered into the analysis. For the first step (one variable in the analysis), Wilks' lambda was calculated using (5.3). For the rest of the analysis (two or more variables in the analysis) Wilks' lambda was calculated using (5.4). The order in which variables were entered into the analysis and the value of Wilks' lambda at each step are recorded in Table 5.2. By noting the sequence in which the variables were entered into the SLDA, the input variables were ranked in order of importance, with the variable entered first considered more important than that entered last. Based on this ranking, variable subsets of increasing size were then created as shown in Table 5.3. Each of these subsets of variables was applied to the DS classifier and its performance was evaluated using OB_{FS} (5.1). The results relating to the performance of each subset are recorded in Table 5.4.

Table 5.2 Order in which the input variables are entered into the SLDA analysis and the associated value of Wilks' lambda, Λ.

Step	Variable entered	Wilks' lambda, Λ
1	v_8	0.282
2	v_{17}	0.199
3	v_2	0.170
4	v_{15}	0.137
5	v_{14}	0.123
6	v_3	0.120
7	v_7	0.114
8	v_{16}	0.111
9	v_1	0.093
10	v_{11}	0.087
11	v_4	0.082
12	v_{18}	0.079
13	v_9	0.078
14	v_5	0.077
15	v_6	0.075
16	v_{13}	0.074
17	v_{10}	0.074
18	v_{12}	0.073

Table 5.3 Subsets of variables created using the SLDA procedure.

Subset	Variable subset
1	$\{v_8\}$
2	$\{v_8, v_{17}\}$
3	$\{v_8, v_{17}, v_2\}$
4	$\{v_8, v_{17}, v_2, v_{15}\}$
5	$\{v_8, v_{17}, v_2, v_{15}, v_{14}\}$
6	$\{v_8, v_{17}, v_2, v_{15}, v_{14}, v_3\}$
7	$\{v_8, v_{17}, v_2, v_{15}, v_{14}, v_3, v_7\}$
8	$\{v_8, v_{17}, v_2, v_{15}, v_{14}, v_3, v_7, v_{16}\}$
9	$\{v_8, v_{17}, v_2, v_{15}, v_{14}, v_3, v_7, v_{16}, v_1\}$
10	$\{v_8, v_{17}, v_2, v_{15}, v_{14}, v_3, v_7, v_{16}, v_1, v_{11}\}$
11	$\{v_8, v_{17}, v_2, v_{15}, v_{14}, v_3, v_7, v_{16}, v_1, v_{11}, v_4\}$
12	$\{v_8, v_{17}, v_2, v_{15}, v_{14}, v_3, v_7, v_{16}, v_1, v_{11}, v_4, v_{18}\}$
13	$\{v_8, v_{17}, v_2, v_{15}, v_{14}, v_3, v_7, v_{16}, v_1, v_{11}, v_4, v_{18}, v_9\}$
14	$\{v_8, v_{17}, v_2, v_{15}, v_{14}, v_3, v_7, v_{16}, v_1, v_{11}, v_4, v_{18}, v_9, v_5\}$
15	$\{v_8, v_{17}, v_2, v_{15}, v_{14}, v_3, v_7, v_{16}, v_1, v_{11}, v_4, v_{18}, v_9, v_5, v_6\}$
16	$\{v_8, v_{17}, v_2, v_{15}, v_{14}, v_3, v_7, v_{16}, v_1, v_{11}, v_4, v_{18}, v_9, v_5, v_6, v_{13}\}$
17	$\{v_8, v_{17}, v_2, v_{15}, v_{14}, v_3, v_7, v_{16}, v_1, v_{11}, v_4, v_{18}, v_9, v_5, v_6, v_{13}, v_{10}\}$
18	$\{v_8, v_{17}, v_2, v_{15}, v_{14}, v_3, v_7, v_{16}, v_1, v_{11}, v_4, v_{18}, v_9, v_5, v_6, v_{13}, v_{10}, v_{12}\}$

Table 5.4 Performance of the subsets of variables selected using the SLDA procedure.

Subset	OB_{FS}	In-sample		Out-of-sample	
		Accuracy,%	OB	Accuracy,%	OB
1	4.3185	96.52	0.8342	97.62	0.8335
2	3.1870	97.62	0.8060	97.62	0.8054
3	4.4023	96.28	0.6857	97.62	0.6843
4	1.6114	99.01	0.6242	100.00	0.6237
5	2.9715	97.62	0.5905	97.62	0.5904
6	2.9379	97.62	0.5569	97.62	0.5567
7	2.9355	97.62	0.5545	97.62	0.5569
8	2.9126	97.62	0.5316	97.62	0.5342
9	3.0052	97.50	0.5081	95.24	0.5125
10	2.8286	97.62	0.4476	97.62	0.4517
11	5.1529	95.24	0.3910	95.24	0.3929
12	3.8622	96.52	0.3779	95.24	0.3832
13	3.9197	96.46	0.3773	92.86	0.3840
14	3.2173	97.15	0.3718	95.24	0.3818
15	2.7401	97.62	0.3591	97.62	0.3697
16	2.7367	97.62	0.3557	97.62	0.3685
17	2.7349	97.62	0.3539	97.62	0.3708
18	2.7173	97.62	0.3363	97.62	0.3541

There is no emerging trend in the behaviour of OB_{FS} as the number of variables in the input set is increased. OB_{FS} reaches a minimum value of 1.6114 when there are four variables in the input set, namely the set $\{v_8, v_{17}, v_2, v_{15}\}$. This is mirrored in the

behaviour of the in-sample accuracy, which shows no particular trend as the number of variables in the input set is increased. The in-sample accuracy reaches a maximum of 99.01% at the same point that OB_{FS} reaches a minimum. In contrast, the in-sample OB shows a decreasing trend as the number of variables in the subset is increased.

The best subset selected by the SLDA procedure - i.e. the subset with the lowest associated OB_{FS} - the subset $\{v_8, v_{17}, v_2, v_{15}\}$, is also the subset with the highest out-of-sample accuracy (100%). As for OB_{FS} and the in-sample accuracy, there is no evident trend in the behaviour of the out-of-sample accuracy as the number of variables in the input set is increased. Similar to the in-sample OB, the out-of-sample OB shows a generally decreasing trend as the number of input variables in the input set is increased.

Finally, for the best subset, a simplex plot showing the simplex coordinates of the final combined BOE (BOE_c) for the out-of-sample subjects is shown in Figure 5.1.

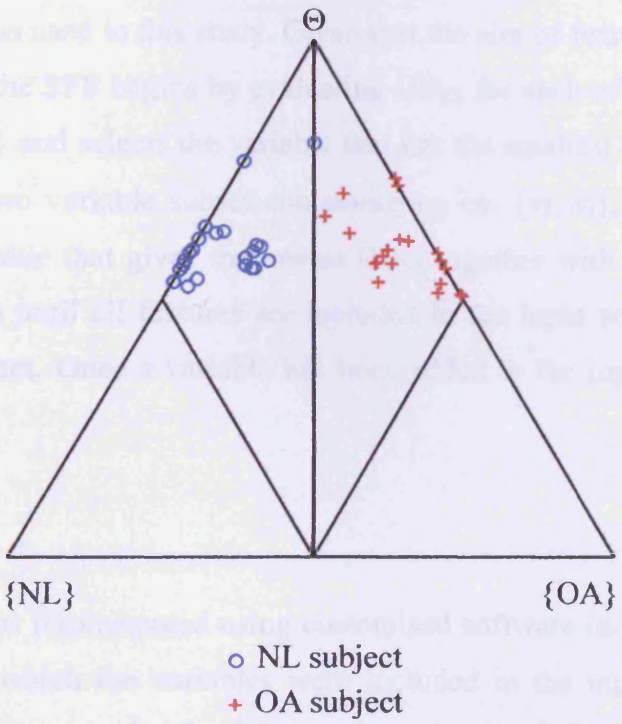


Figure 5.1 Simplex plot showing simplex coordinates of the combined BOE (BOE_c) for the out-of-sample subjects using subset 4, $\{v_8, v_{17}, v_2, v_{15}\}$.

The simplex coordinates of all subjects lie on the correct sides of the decision boundary. However, the simplex coordinates of all the subjects lie within the non-dominant areas of the simplex plot. This implies that the level of uncertainty associated with each combined body of evidence, $m_c(\Theta)$ is high. Indeed for each subject $m_c(\Theta) > m_c(\{OA\}) + m_c(\{NL\})$. This suggests that there is a low level of certainty in the correct classification of the subjects to their respective groups.

5.4 SEQUENTIAL SELECTION METHODS (SSM)

5.4.1 SSM Theory

There are two types of SSM: sequential forward selection (SFS) and sequential backward selection (SBS). SFS begins with the empty set and iteratively adds features until some stopping criterion is met. Conversely, SBS starts with the full set of features and iteratively removes features until some stopping criterion is met (Jain and Zongker, 1997). The SFS is preferential to the SBS since it is more costly to evaluate an objective function for large sets than for small ones (Jain and Zongker, 1997). Consequently, SFS is used in this study. Given that the aim of feature selection here is to minimise OB_{FS}, the SFS begins by evaluating OB_{FS} for each of the variables in the set $V = \{v_1, v_2, ...v_p\}$ and selects the variable that has the smallest OB_{FS}, say v_3. Then, the OB_{FS} of each two variable subset containing v_3, i.e. $\{v_3, v_i\}$, $i = 1: p$ $(p \neq 3)$ is evaluated. The variable that gives the lowest OB_{FS} together with v_3 is retained. This procedure continues until all features are included in the input set or some stopping criterion has been met. Once a variable has been added to the input set, it cannot be removed.

5.4.2 SFS Results

A SFS procedure was implemented using customised software in Matlab 5.3 (Matlab Inc.). The order in which the variables were included in the input set is shown in Table 5.5. The performance of each of these subsets is recorded in Table 5.6.

Table 5.5 Subsets of variables created using the SFS method.

Subset	Variable subset
1	$\{v_8\}$
2	$\{v_8, v_{16}\}$
3	$\{v_8, v_{16}, v_2\}$
4	$\{v_8, v_{16}, v_2, v_{17}\}$
5	$\{v_8, v_{16}, v_2, v_{17}, v_5\}$
6	$\{v_8, v_{16}, v_2, v_{17}, v_5, v_{14}\}$
7	$\{v_8, v_{16}, v_2, v_{17}, v_5, v_{14}, v_6\}$
8	$\{v_8, v_{16}, v_2, v_{17}, v_5, v_{14}, v_6, v_{11}\}$
9	$\{v_8, v_{16}, v_2, v_{17}, v_5, v_{14}, v_6, v_{11}, v_{13}\}$
10	$\{v_8, v_{16}, v_2, v_{17}, v_5, v_{14}, v_6, v_{11}, v_{13}, v_{10}\}$
11	$\{v_8, v_{16}, v_2, v_{17}, v_5, v_{14}, v_6, v_{11}, v_{13}, v_{10}, v_7\}$
12	$\{v_8, v_{16}, v_2, v_{17}, v_5, v_{14}, v_6, v_{11}, v_{13}, v_{10}, v_7, v_9\}$
13	$\{v_8, v_{16}, v_2, v_{17}, v_5, v_{14}, v_6, v_{11}, v_{13}, v_{10}, v_7, v_9, v_1\}$
14	$\{v_8, v_{16}, v_2, v_{17}, v_5, v_{14}, v_6, v_{11}, v_{13}, v_{10}, v_7, v_9, v_1, v_3\}$
15	$\{v_8, v_{16}, v_2, v_{17}, v_5, v_{14}, v_6, v_{11}, v_{13}, v_{10}, v_7, v_9, v_1, v_3, v_{18}\}$
16	$\{v_8, v_{16}, v_2, v_{17}, v_5, v_{14}, v_6, v_{11}, v_{13}, v_{10}, v_7, v_9, v_1, v_3, v_{18}, v_{15}\}$
17	$\{v_8, v_{16}, v_2, v_{17}, v_5, v_{14}, v_6, v_{11}, v_{13}, v_{10}, v_7, v_9, v_1, v_3, v_{18}, v_{15}, v_4\}$
18	$\{v_8, v_{16}, v_2, v_{17}, v_5, v_{14}, v_6, v_{11}, v_{13}, v_{10}, v_7, v_9, v_1, v_3, v_{18}, v_{15}, v_4, v_{12}\}$

Table 5.6 Performance of the subsets of variables created using the SFS method.

Subset	OB_{FS}	In-sample		Out-of-sample	
		Accuracy,%	OB	Accuracy,%	OB
1	4.3185	96.52	0.8342	97.62	0.8335
2	3.1756	97.62	0.7946	97.62	0.7937
3	2.3018	98.67	0.6758	100.00	0.6744
4	0.6547	100.00	0.6547	100.00	0.6535
5	0.6411	100.00	0.6411	100.00	0.6441
6	0.6051	100.00	0.6051	100.00	0.6086
7	0.5793	100.00	0.5793	100.00	0.5832
8	0.5076	100.00	0.5076	100.00	0.5112
9	0.5008	100.00	0.5008	100.00	0.5066
10	0.4971	100.00	0.4971	100.00	0.5070
11	0.4952	100.00	0.4952	100.00	0.5077
12	0.5523	99.94	0.4942	100.00	0.5081
13	0.8216	99.65	0.4732	100.00	0.4888
14	2.0749	98.37	0.4489	95.24	0.4641
15	2.7578	97.68	0.4349	95.24	0.4516
16	2.786	97.62	0.4050	95.24	0.4227
17	2.7349	97.62	0.3539	97.62	0.3708
18	2.7173	97.62	0.3363	97.62	0.3541

The value of OB_{FS} decreases as the number of variables in the input set is increased, reaches a minimum, increases again and finally decreases slightly. OB_{FS} reaches a minimum value of 0.4952 when there are 11 variables in the input set, namely the subset $\{v_8, v_{16}, v_2, v_{17}, v_5, v_{14}, v_6, v_{11}, v_{13}, v_{10}, v_7\}$. As the number of variables in the

input set is increased, the in-sample accuracy increases until it reaches a maximum of 100%, remains at this maximum and then begins to decrease until it reaches a plateau of 97.62%. The in-sample OB shows a decreasing trend as the number of variables in the subset is increased.

The best subset selected by the SFS procedure, the subset {v_8, v_{16}, v_2, v_{17}, v_5, v_{14}, v_6, v_{11}, v_{13}, v_{10}, v_7}, is one of the subsets with the highest out-of-sample accuracy (100%). Similar to the in-sample OB, the out-of-sample OB shows a decreasing trend as the number of input variables in the input set is increased.

Finally, for the best subset, a simplex plot showing the simplex coordinates of the BOE_c for the out-of-sample subjects is shown in Figure 5.2.

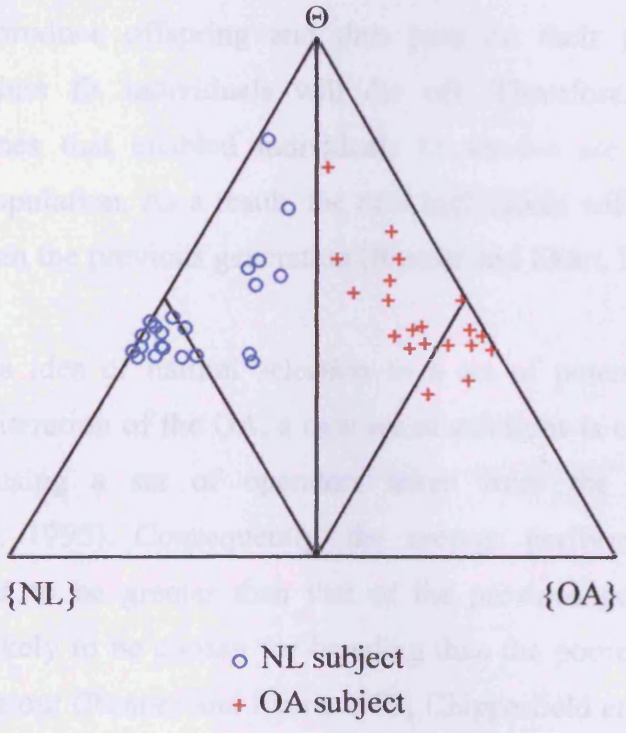


Figure 5.2 Simplex plot showing simplex coordinates of the combined BOE (BOE_c) for the out-of-sample subjects using variable subset 11, {v_8, v_{16}, v_2, v_{17}, v_5, v_{14}, v_6, v_{11}, v_{13}, v_{10}, v_7}.

The simplex coordinates of all out-of-sample subjects lie on the correct sides of the decision boundary. The simplex coordinates of thirteen NL and seven OA subjects lie within the dominant regions of the simplex-plot. The simplex coordinates of the remainder of subjects lie within the non-dominant regions. For those that lie within the non-dominant regions of the simplex plot, the level of certainty in their correct classification is lower than for those that lie within the dominant regions.

5.5 GENETIC ALGORITHMS (GA)

5.5.1 GA Theory

Like the Simulated Annealing algorithm (as described in Chapter 3, section 3.5.2), the GA is a search technique that mimics a naturally occurring process - the biological process of natural selection (Brill *et al.*, 1992). Natural selection, or survival of the fittest, implies that individuals that are more suited to their environment are more likely to survive, produce offspring and thus pass on their genes to the next generation, whilst less fit individuals will die off. Therefore, over subsequent generations, the genes that enabled individuals to survive are likely to become widespread in the population. As a result, the new individuals will be more suited to their environment than the previous generation (Renner and Ekárt, 2003).

The GA applies this idea of natural selection to a set of potential solutions to a problem. During an iteration of the GA, a new set of solutions is created by breeding existing solutions using a set of operators taken from the field of genetics (Chipperfield *et al.*, 1995). Consequently, the average performance of the new solutions is expected to be greater than that of the previous generation, as good solutions are more likely to be chosen for breeding than the poorer solutions, which are more likely to die out (Renner and Ekárt, 2003; Chipperfield *et al.*, 1995). Unlike most other feature selection methods that search for a single solution, the GA is a parallel search technique and thus maintains a collection of potential solutions (Renner and Ekárt, 2003; Chipperfield *et al.*, 1994). Siedlecki and Sklansky (1989) were the first to introduce the GA as a feature selection method. The application of the GA to feature selection is summarised in Figure 5.3 and is now discussed in more detail.

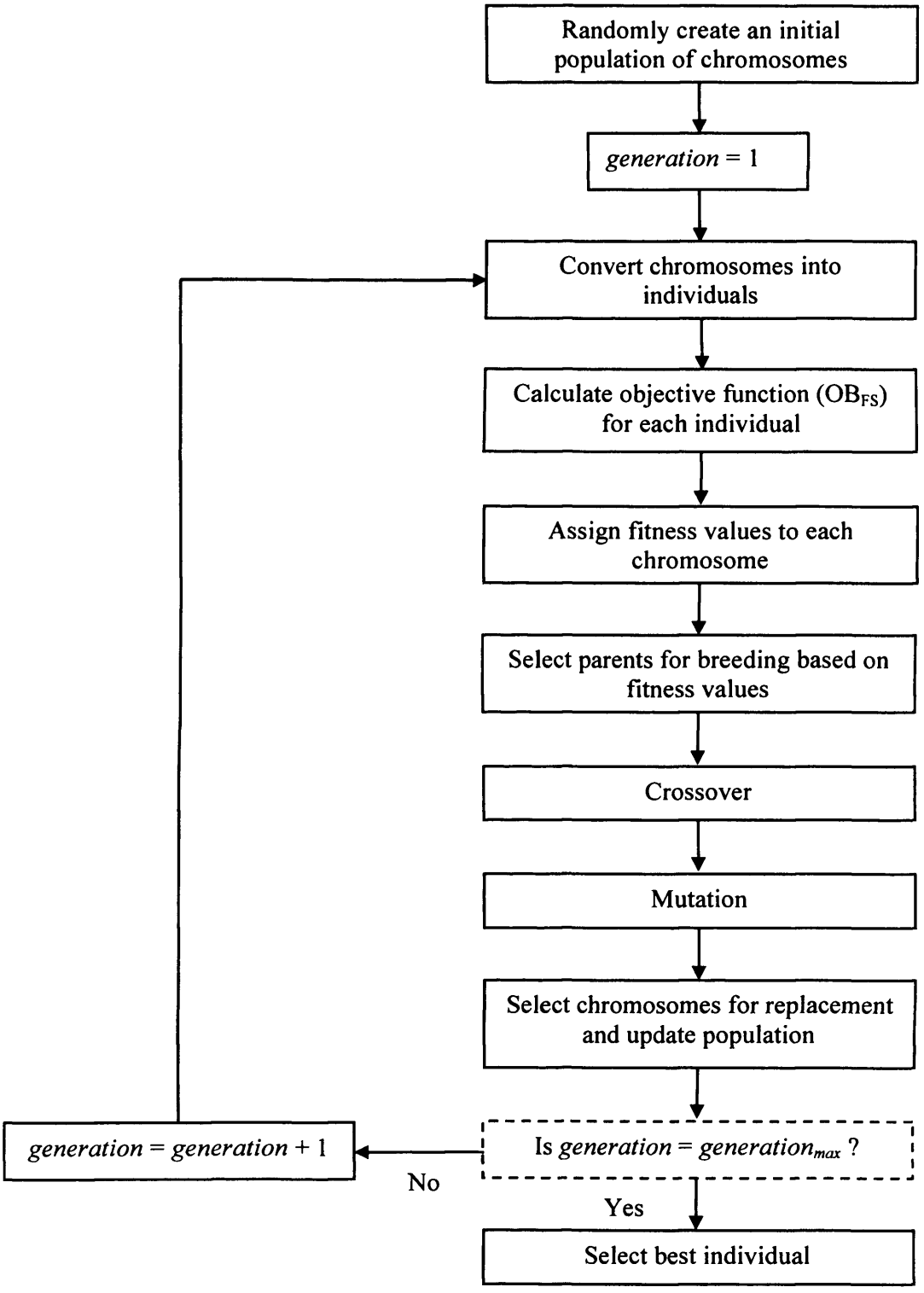


Figure 5.3 Feature selection using a GA.

Let the set of features (input variables) be $V = \{v_1, v_2, ..., v_p\}$. A potential solution to the problem of feature selection, an ***individual***, is any subset of V. In the GA, an individual is represented by a finite sequence of zeros or ones. This representation is called a ***chromosome***. Although other coding techniques are available, binary coding is the most commonly used (Chipperfield *et al.*, 1994) and is the representation employed by Siedlecki and Sklansky (1989). In the case of feature selection, the zero or one in a binary coded chromosome implies the respective absence or presence of a variable in the subset (Siedlecki and Sklansky, 1989). The variables that constitute the chromosome are defined as ***genes***. An example of a chromosome is given in Figure 5.4.

The GA begins by randomly creating a finite number of chromosomes, m. Collectively these chromosomes are defined as a ***population***. The size of the population, m remains constant throughout the optimisation process. In this study, an initial population is created by assigning random binary values to m chromosomes of length p. Subsequently the chromosomes are decoded into ***individuals***. An example of this decoding process is given in Figure 5.5.

Following this transformation process, the OB_{FS} of each individual (and its associated chromosome) is calculated (see section 5.2). Subsequently the performance, or ***fitness***, of each chromosome can be assessed. The GA ranks the chromosomes in descending order according to their OB_{FS}; the least fit chromosome, with a large associated OB_{FS}, is given a ranking of one whilst the fittest chromosome is given a ranking of m. Each chromosome, w_j, is subsequently assigned a fitness value, $f(w_j)$, according to its ranking:

$$f(w_j) = 2 - s + \frac{2 \times (s-1) \times (r_j - 1)}{(m-1)} \tag{5.5}$$

where s is the selected pressure (the probability of the best chromosome being selected compared to the average probability of selection of all chromosomes), r_j is the ranking of the chromosome w_j and m the size of the population (Chipperfield *et al.*, 1994). For $s = 2$, equation (5.5) reduces to

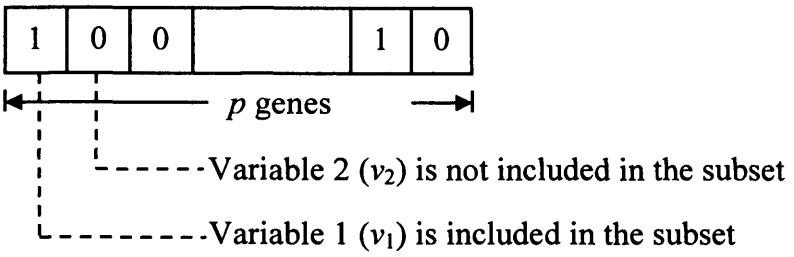


Figure 5.4 A binary coded chromosome. The gene values of zero or one imply the respective absence or presence of a variable in an individual. (Taken from Raymer *et al.*, 2000, pp.166).

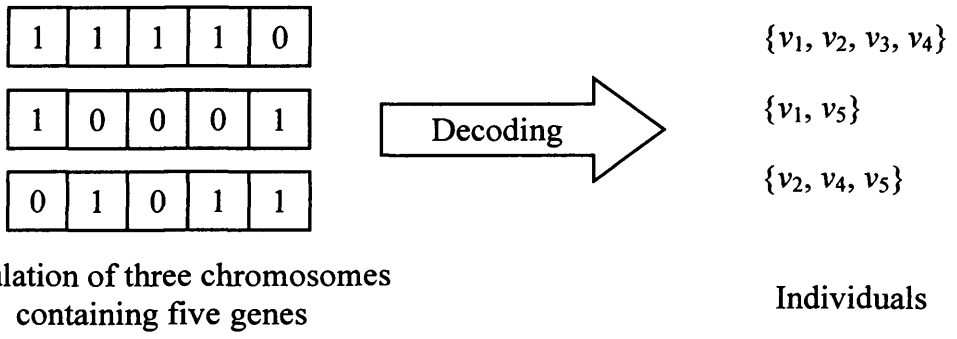


Figure 5.5 Decoding population of chromosomes into individuals (subsets of variables).

$$f(w_j) = \frac{2 \times (r_j - 1)}{(m - 1)} \tag{5.6}$$

An example of ranking is given in Table 5.7. Six chromosomes, w_j, (j = 1:6) are ranked according to their OB in descending order. Subsequently the fitness values, $f(w_j)$ (j = 1:6) are calculated using equation (5.6).

Table 5.7 An example of ranking of chromosomes and assignment of fitness values.

Chromosome, w_j	**OB**	**Ranking, r_j**	**Fitness, $f(w_j)$**
w_1	0.234	4	= 2 × (4 - 1) / (6 –1) = 1.2
w_2	0.451	2	= 2 × (2 - 1) / (6 –1) = 0.4
w_3	0.632	1	= 2 × (1 - 1) / (6 –1) = 0
w_4	0.133	6	= 2 × (6 - 1) / (6 –1) = 2
w_5	0.178	5	= 2 × (4 - 1) / (6 –1) = 1.6
w_6	0.369	3	= 2 × (4 - 1) / (6 –1) = 0.8

Subsequently a number of new chromosomes are produced according to three genetic operators: ***selection***, ***crossover*** and ***mutation***. These operators enable genetic information to be exchanged between chromosomes. The ratio of the number of new chromosomes produced to the size of the original population is termed the ***generation gap***.

Based on the computed fitness values, a predetermined number of chromosomes, n_{sel}, are ***selected*** from the population for breeding using stochastic universal sampling (Baker, 1987). The manner in which the chromosomes are selected is now explained using the example where four individuals are selected for breeding from the population of six individuals given in Table 5.7.

The chromosomes are arranged in a random order as shown in Table 5.8 and a random number, p_{sel}, is generated in the range $[0, q_i]$ where

$$q_j = \left(\frac{\sum_{j=1}^{m} f(w_j)}{n_{sel}} \right) \tag{5.7}$$

Table 5.8 Calculation of fitness interval.

Chromosome, w_j	Fitness $f(w_j)$	Cumulative Fitness	Fitness interval
w_2	0.4	0.4	[0, 0.4]
w_1	1.2	1.6	[0.4,1.6]
w_6	0.8	2.4	[1.6,2.4]
w_3	0	2.4	-
w_5	1.6	4	[2.4,4]
w_4	2	6	[4,6]

Chromosomes are subsequently chosen for breeding by generating n_{sel} equally spaced pointers, $[p_{sel}, (p_{sel} + q_j), \ldots, (p_{sel} + q_j(n_{sel} - 1))]$.

In the previous example, $n_{sel} = 4$ and $\sum_{j=1}^{6} f(w_j) = 6$ (see Table 5.8). From equation (5.7) this gives $q_j = 1.5$. Let p_{sel} be a random number selected in the range [0, 1.5], say 0.15. Then the four pointers are [0.15, (0.15 + 1.5), (0.15 + (1.5 × 2)), (0.15 + (1.5 × 3))] i.e. [0.15, 1.65, 3.15, 4.65]. These pointers are to be found within the fitness intervals of chromosomes w_2, w_6, w_5 and w_4 respectively as shown in Table 5.8. Thus, chromosomes w_2, w_6, w_5 and w_4 are chosen for breeding.

The probability of a chromosome w_j being selected, $F(w_j)$ is related to its fitness relative to the whole population, and is given by:

$$F(w_j) = \frac{f(w_j)}{\sum_{j=1}^{m} f(w_j)} \tag{5.8}$$

where $f(w_j)$ is the fitness of the chromosome w_j (Chipperfield *et al.*, 1994). This shows that selection is biased towards more fit chromosomes (Chipperfield *et al.*, 1994), i.e. that chromosomes with a high level of fitness relative to the whole population are more likely to be selected for breeding and vice versa.

The selected chromosomes are subsequently paired up for crossover. The individuals selected for breeding are called ***parents*** and the new individuals produced during breeding, ***offspring***. Crossover produces offspring that contain some part of both their parents' genetic information. The simplest crossover operator is single-point crossover as illustrated in Figure 5.6. Each chromosome pair is crossed over at a randomly selected crossover point according to some probability p_c, defined the ***crossover probability***. For a chromosome of length p, the crossover point is selected randomly from the genes 1 to $p - 1$.

Following crossover, the mutation operator is applied to the offspring as shown in Figure 5.7. Mutation changes the value of each gene from a zero to a one (or a one to a zero) according to the probability p_m, defined the ***mutation probability***. The mutation probability increases the variation in the population and reduces the risk of the GA converging to a poor solution (i.e. to a local rather than the global minimum) (Renner and Ekárt, 2003; Siedlecki and Sklansky, 1989; Chipperfield *et al.*, 1994).

Subsequent to the production of offspring, the current population is updated. When using a generation gap of less than one the offspring must be inserted into the current population to maintain its size. Following the work of Su and Wu (2000), fitness-based reinsertion is utilised to select which members of the current population should be replaced.

An illustration of this fitness-based reinsertion is given using the earlier example, where the parents w_2, w_4, w_5 and w_6 were selected for breeding. Let the offspring produced during crossover and mutation be w_7 and w_8. Using fitness-based reinsertion, these two offspring will be reinserted into the current population to replace the two least fit chromosomes. Table 5.7 shows that the two least fit chromosomes are w_2 and w_3. Thus, the new updated population becomes $\{w_1, w_4, w_5, w_6, w_7, w_8\}$.

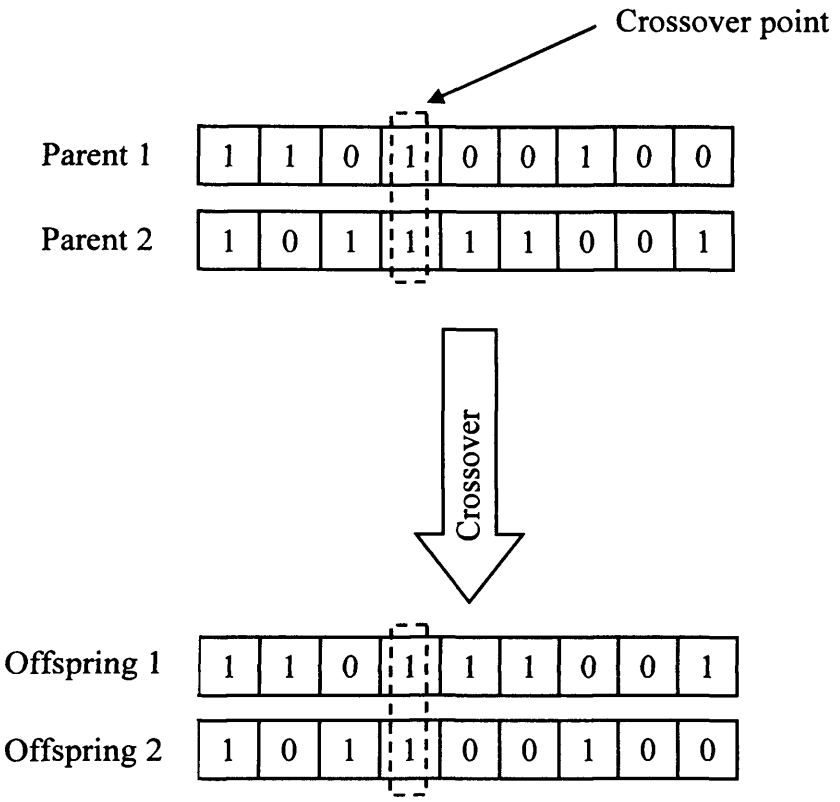


Figure 5.6 The single point crossover operator is applied to two parent chromosomes, Parent 1 and Parent 2, to produce two offspring, Offspring 1 and Offspring 2, which contain a mixture of their parents' genetic information.

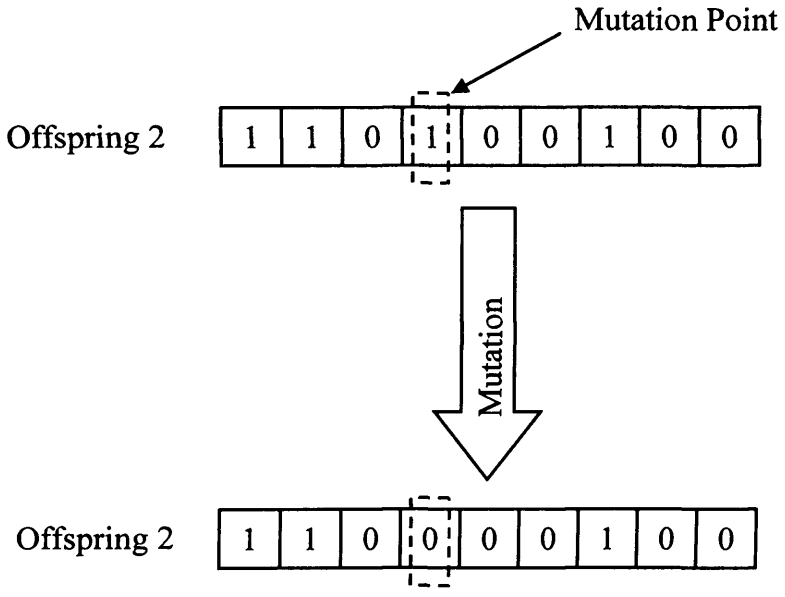


Figure 5.7 The mutation operator is applied to Offspring 2. A randomly selected gene, as shown by the dotted line, is changed from a one to a zero.

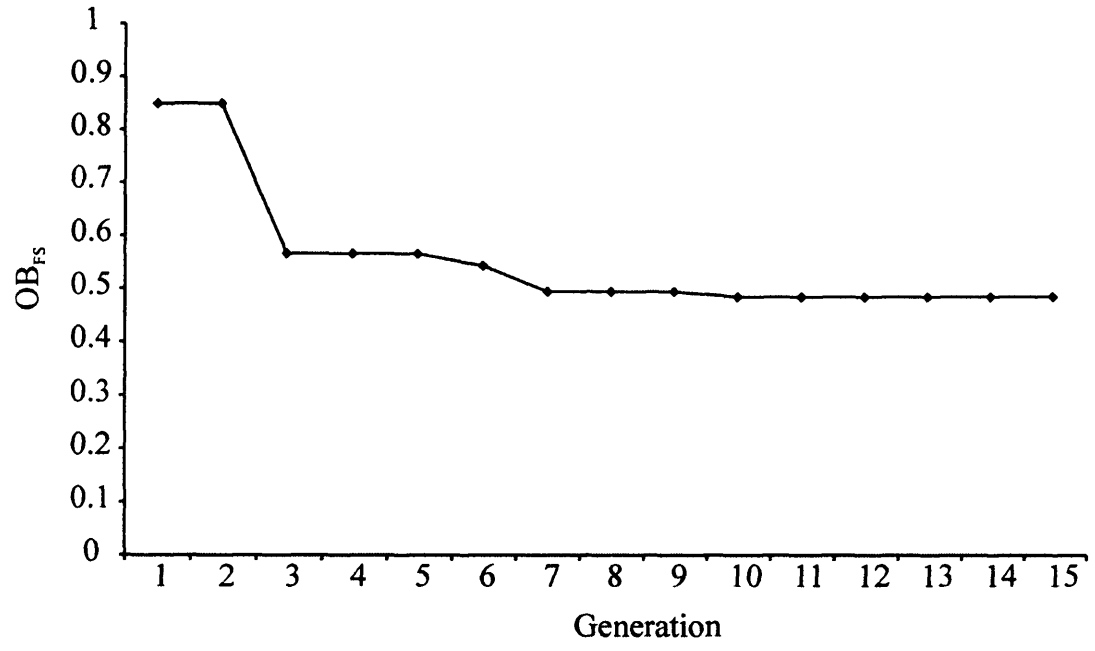
The cycle of selection, crossover, mutation and reinsertion is implemented repeatedly until a predetermined number of iterations (***generations***) is reached. Finally, the best individual, i.e. the individual with the lowest associated OB_{FS}, is selected from the final population.

5.5.2 GA Results

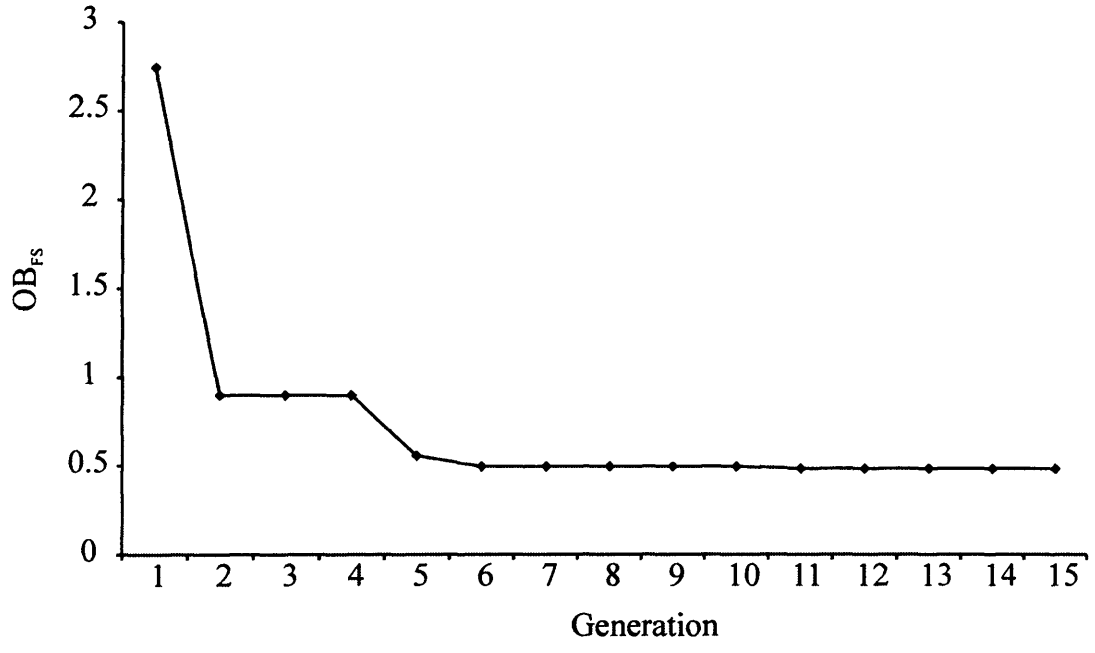
A customised GA was programmed using Matlab 5.3 (Matlab Inc.). On examination of the work of Su and Wu (2000) the following parameters were utilised: a population size of 200, a generation gap of 0.8, a crossover probability of 0.8, a mutation probability of 0.001 and a maximum number of generations set at 15. Ten runs of the algorithm were performed because of its random nature. The best individual selected from the final population for each of the ten runs is recorded in Table 5.9. An example of the way in which the GA converges is given in Figure 5.8 for runs 1 and 2. Although both of these runs converged to the same final solution (Table 5.9), the way in the algorithm converges is different for each run.

Table 5.9 Best individual selected from the final population.

Run	Best Individual
1	$\{v_2, v_5, v_6, v_7, v_8, v_{11}, v_{13}, v_{15}, v_{16}, v_{17}\}$
2	$\{v_2, v_5, v_6, v_7, v_8, v_{11}, v_{13}, v_{15}, v_{16}, v_{17}\}$
3	$\{v_2, v_5, v_6, v_7, v_8, v_{11}, v_{13}, v_{15}, v_{16}, v_{17}\}$
4	$\{v_2, v_5, v_6, v_7, v_8, v_{10}, v_{11}, v_{13}, v_{14}, v_{16}, v_{17}\}$
5	$\{v_2, v_5, v_6, v_7, v_8, v_{10}, v_{11}, v_{13}, v_{14}, v_{16}, v_{17}\}$
6	$\{v_2, v_5, v_6, v_7, v_8, v_{10}, v_{11}, v_{13}, v_{14}, v_{16}, v_{17}\}$
7	$\{v_2, v_5, v_6, v_7, v_8, v_{11}, v_{13}, v_{15}, v_{16}, v_{17}\}$
8	$\{v_2, v_5, v_6, v_7, v_8, v_{11}, v_{13}, v_{15}, v_{16}, v_{17}\}$
9	$\{v_2, v_5, v_6, v_7, v_8, v_{11}, v_{13}, v_{15}, v_{16}, v_{17}\}$
10	$\{v_2, v_5, v_6, v_7, v_8, v_{11}, v_{13}, v_{15}, v_{16}, v_{17}\}$



(a)



(b)

Figure 5.8 Typical runs of the GA algorithm: (a) run 1 (b) run 2. The random nature of the algorithm means that each run may converge differently.

In each run the algorithm converged to one of two similar solutions, $\{v_2, v_5, v_6, v_7, v_8, v_{11}, v_{13}, v_{15}, v_{16}, v_{17}\}$ (subset 1) and $\{v_2, v_5, v_6, v_7, v_8, v_{10}, v_{11}, v_{13}, v_{14}, v_{16}, v_{17}\}$ (subset 2). The performance of these two subsets is shown in Table 5.10. Both of these subsets produced an in-sample and out-of-sample accuracy of 100%. However, subset 1 has a slightly lower associated OB_{FS} than subset 2 due to a lower in-sample OB and thus is chosen as the best subset.

Table 5.10 Performance of subset 1 (runs 1, 2, 3, 7, 8, 9 and 10) and subset 2 (runs 4, 5 and 6).

Subset	OB_{FS}	In-sample		Out-of-sample	
		Accuracy	OB	Accuracy	OB
1	0.4855	100.00	0.4855	100.00	0.4940
2	0.4952	100.00	0.4952	100.00	0.5077

For subset 1, a simplex plot showing the simplex coordinates of the BOE_c for the out-of-sample subjects is shown in Figure 5.9.

The simplex coordinates of all the subjects lie on the correct sides of the decision boundary. The simplex coordinates of ten NL subjects and nine OA subjects lie within the dominant regions of the simplex plot. For the ten NL subjects this implies that $m_c(\{NL\}) > m_c(\{OA\}) + m_c(\Theta)$. Similarly for the nine OA subjects $m_c(\{OA\}) > m_c(\{NL\}) + m_c(\Theta)$. However, for the remainder of the subjects, their simplex coordinates lie within the non-dominant regions of the simplex plot.

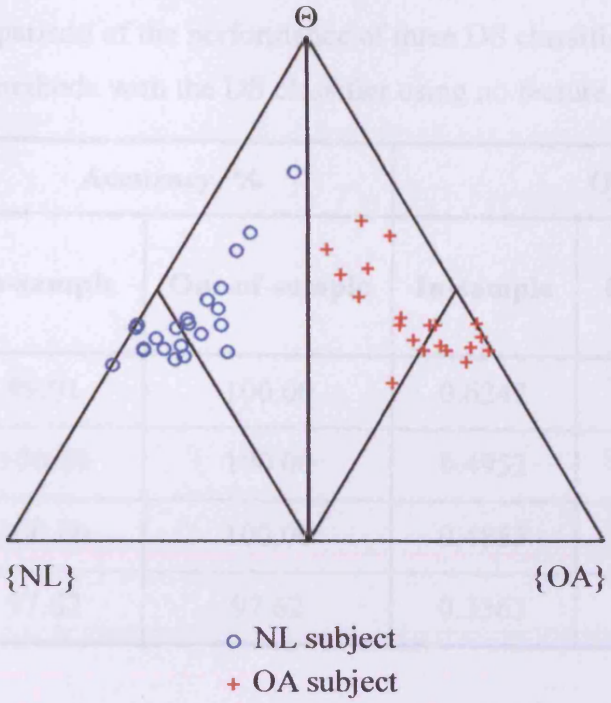


Figure 5.9 Simplex plot showing simplex coordinates of the combined BOE (BOE_c) for the out-of-sample subjects using subset 1, $\{v_2, v_5, v_6, v_7, v_8, v_{11}, v_{13}, v_{15}, v_{16}, v_{17}\}$.

5.6 DISCUSSION OF RESULTS

This study was conducted to examine the effect of using a feature selection algorithm on the performance of the DS classifier. Three feature selection methods, namely SLDA, SFS and GA, have been described in detail and used in conjunction with the DS classifier outlined in Chapter 3. The results from this study have been presented and will now be discussed.

The in-sample and out-of-sample accuracy and OB results are summarised in Table 5.11. The results in terms of the positioning of the simplex coordinates of the out-of-sample subjects are depicted in Figure 5.10.

Table 5.11 Comparison of the performance of three DS classifiers using feature selection methods with the DS classifier using no feature selection.

Feature Selection Method	Accuracy, %		OB	
	In-sample	Out-of-sample	In-sample	Out-of-sample
SLDA	99.01	100.00	0.6242	0.6237
SFS	100.00	100.00	0.4952	0.5077
GA	100.00	100.00	0.4855	0.4940
None	97.62	97.62	0.3363	0.3541

The in-sample and out-of-sample classification accuracy and in-sample and out-of-sample OB results of the three DS classifiers using feature selection are compared to the DS classifier that does not use feature selection (see Chapter 4, section 4.1.2.1 test 1Ii). Using feature selection produces a superior classifier in terms of the in-sample and out-of-sample classification accuracy. The DS classifiers that used feature selection produced out-of-sample accuracies of 100% compared to the 97.62% of the DS classifier with no feature selection. All of the classifiers with feature selection produced higher in-sample accuracies than the DS classifier with no feature selection. The DS classifiers with SFS and GA produced in-sample accuracies of 100%, whilst the DS classifier with SLDA produced a slightly lower in-sample accuracy of 99.01%.

However, the use of feature selection also results in an increase in the in-sample and out-of-sample OB. This indicates that there is an increase in the level of associated uncertainty in the classification. This is particularly evident in Figure 5.10, which shows a comparison of the simplex coordinates of the out-of-sample subjects for the DS classifier with no feature selection (Figure 5.10a) with the three DS classifiers using feature selection (Figures 5.10b, 5.10c and 5.10d). In general, in Figures 5.10b, 5.10c and 5.10d the distance of the simplex coordinates from the uncertainty vertex is much less than in Figure 5.10a.

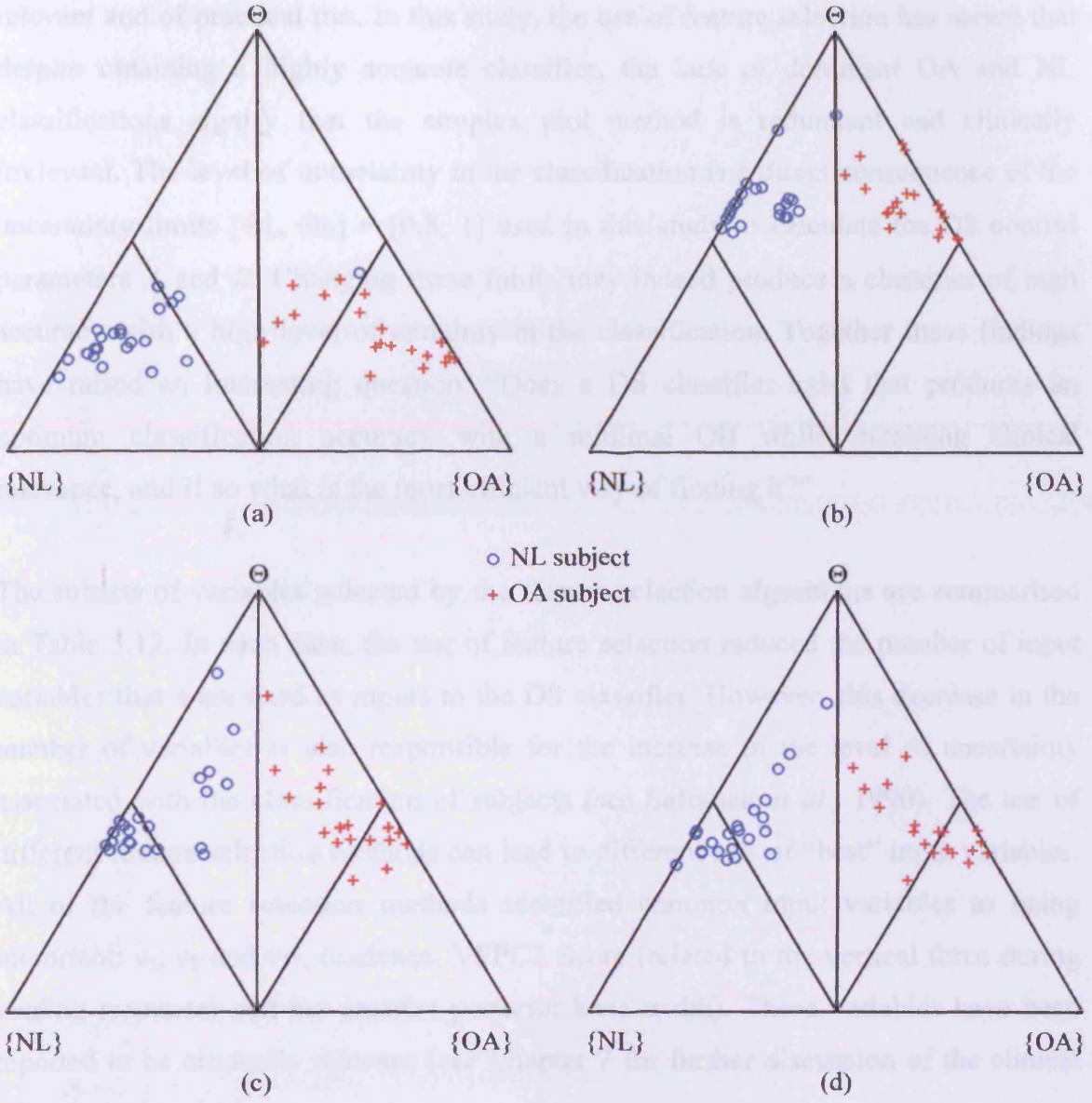


Figure 5.10 Simplex plots showing the simplex coordinates of the BOE_cs for the out-of-sample subjects for (a) DS classifier with no feature selection, (b) DS classifier with SLDA, (c) DS classifier with SFS and (d) DS classifier with GA.

The use of the simplex plot highlights that although obtaining a high classification accuracy is of first importance, it is necessary to ensure that the method is still relevant and of practical use. In this study, the use of feature selection has meant that despite obtaining a highly accurate classifier, the lack of dominant OA and NL classifications signify that the simplex plot method is redundant and clinically irrelevant. The level of uncertainty in the classification is a direct consequence of the uncertainty limits $[\Theta_L, \Theta_U] = [0.8, 1]$ used in this study to calculate the DS control parameters *A* and *B*. Changing these limits may indeed produce a classifier of high accuracy with a high level of certainty in the classification. Together these findings have raised an interesting question: "Does a DS classifier exist that produces an optimum classification accuracy with a minimal OB whilst retaining clinical relevance, and if so what is the most efficient way of finding it?"

The subsets of variables selected by the feature selection algorithms are summarised in Table 5.12. In each case, the use of feature selection reduced the number of input variables that were used as inputs to the DS classifier. However, this decrease in the number of variables is also responsible for the increase in the level of uncertainty associated with the classification of subjects (see Safranek *et al.*, 1990). The use of different feature selection methods can lead to different sets of "best" input variables. All of the feature selection methods identified common input variables as being important: v_2, v_8 and v_{17}, (cadence, VFPC2 Score (related to the vertical force during loading response) and the anterior-posterior knee width). These variables have been reported to be clinically relevant (see Chapter 7 for further discussion of the clinical relevance of variables).

Table 5.12 Subsets of variables selected by the different feature selection methods.

Method	Subset of variables
SLDA	$\{v_2, v_8, v_{15}, v_{17}\}$
SFS	$\{v_2, v_5, v_6, v_7, v_8, v_{10}, v_{11}, v_{13}, v_{14}, v_{16}, v_{17}\}$
GA	$\{v_2, v_5, v_6, v_7, v_8, v_{11}, v_{13}, v_{15}, v_{16}, v_{17}\}$

Of the three feature selection methods, the DS classifier with GA produced the highest classification accuracy along with the lowest OB. However, this method is much more computationally intensive than the SFS method, which produces the same accuracy and only a slightly larger OB.

This chapter has investigated the effect of using automatic feature selection on the performance of the DS classifier. The results for three different feature selection methods have been presented and discussed. Conclusions based on this study and suggestions for further work are given in Chapter 8.

CHAPTER 6

TOTAL KNEE REPLACEMENT STUDY

This chapter investigates the novel application of the Dempster-Shafer (DS) classifier as a tool for assessing the outcome of total knee replacement (TKR) surgery. The chapter begins by providing a brief background to the study. The results of the study are presented and discussed for a set of nine patients followed by a discussion of the overall results.

6.1 INTRODUCTION

TKR surgery is used to treat approximately 35,000 patients with knee osteoarthritis (OA) in the United Kingdom each year (Moran and Horton, 2000). During surgery, the degenerative cartilaginous surfaces of the joint are replaced with the articulating surfaces of prosthetic components. The procedure is performed primarily to reduce the pain associated with the degenerative disease and to restore a degree of normal (NL) knee function (Andriacchi, 1993; Myles *et al.*, 2002). The beneficial implications of a return towards NL function include better mobility and an improved functional effect on other lower limb joints by removing the need for compensatory gait mechanisms. Additionally, for congruent mobile bearing knees, a return to NL function can result in a decrease in the contact and shear stresses on the articulating surfaces of the replacement joint, thus reducing polyethylene wear and subsequent implant failure.

The need for a universal tool to assess the outcome of TKR surgery is widely recognised. Such a tool should fulfil the following main requirements: (Davies, 2002)

1. Enable direct comparison between subjects
2. Establish the level of benefit achieved by surgery
3. Enable direct comparison between different surgical techniques or implants
4. Use important measurable characteristics of the knee that are clinical variables and are easily quantified

5. Relate the outcome to the clinical results
6. Simplicity

In response to this need, there has been an emergence of two types of systems to assess knee function before and after TKR surgery: the patient-reported scoring systems and gait analysis studies. In general, patient-reported scoring systems attempt to measure patient wellbeing in terms of pain and daily life activities. Examples of such systems are The Knee Outcome Survey (Irrgang *et al.*, 1998); WOMAC (see Davies, 2002); Oxford Knee (see Davies, 2002). However, they do not offer an objective assessment of the function of the knee. In contrast, gait analysis studies can provide an objective measure of knee function (e.g. Catani *et al.*, 2003; Benedetti *et al.*, 2003; Myles *et al.*, 2002; Fuchs *et al.*, 2002; Whittle and Jefferson, 1989; Chao *et al.*, 1980; Andriacchi, 1993).

This chapter investigates the potential of the DS classifier as an objective tool for assessing the outcome of TKR surgery. Nine patients with knee OA (the TKR sample) were followed before and at three stages after TKR surgery. The study was limited to this size sample because only nine patients had completed a full set of four visits (one pre-operative and three post-operative visits). During each visit, 18 variables relating to their knee function were collated using the methods described in Chapter 2. A DS classifier was trained using the variables of the combined OA and NL sample. The DS control variables k, θ, A and B, were calculated from the variables of the combined OA and NL sample group using the non-optimisation method described in Section 3.5.1. k was calculated using the correlation coefficient method and A and B from the limits $[\Theta_L, \Theta_U] = [0.8, 1]$. These control variables were then used to transform the input variables of the TKR sample into a combined body of evidence (BOE_c) for each patient.

For comparison purposes, at each visit patients completed the Knee Outcome Survey (KOS) as described in Section 2.1.7 and Appendix A and an orthopaedic surgeon conducted a Blinded Observational Analysis (BOA). The results of this study are now presented and discussed for each patient in turn.

6.2 PATIENT 1 (P1)

P1 was followed before and at three stages after TKR surgery. The timing of these visits is given in Table 6.1.

Table 6.1 P1 visit summary.

Visit	Visit type
1	Preoperative
2	3 months postoperative
3	6 months postoperative
4	12 months postoperative

6.2.1 P1 DS classifier results

The BOE_c values for the four visits of P1 are recorded in Table 6.2. The simplex representations of the BOE_cs are depicted in Figure 6.1.

Table 6.2 BOE_c values for the four visits of P1.

BOE_c	Visit			
	1	2	3	4
$m_c(\{OA\})$	0.5469	0.4011	0.5955	0.5489
$m_c(\{NL\})$	0.1366	0.3163	0.0690	0.1718
$m_c(\Theta)$	0.3164	0.2827	0.3354	0.2793

At visit 1 P1 has a dominant OA classification since $m_c(\{OA\}) > [m_c(\{NL\}) + m_c(\Theta)]$. This is reflected in the positioning of its simplex coordinate within the dominant OA classification region of the simplex plot.

At visit 2 $m_c(\{OA\})$ is still the greatest of the three belief values. However, from visit 1 to visit 2 $m_c(\{OA\})$ has decreased to the extent that $m_c(\{OA\}) < [m_c(\{NL\}) + m_c(\Theta)]$. Additionally, there has been an increase in $m_c(\{NL\})$. This change is evident

in Figure 6.1. From visit 1 to visit 2 the simplex coordinate has moved from the dominant OA region to the non-dominant OA region of the simplex plot.

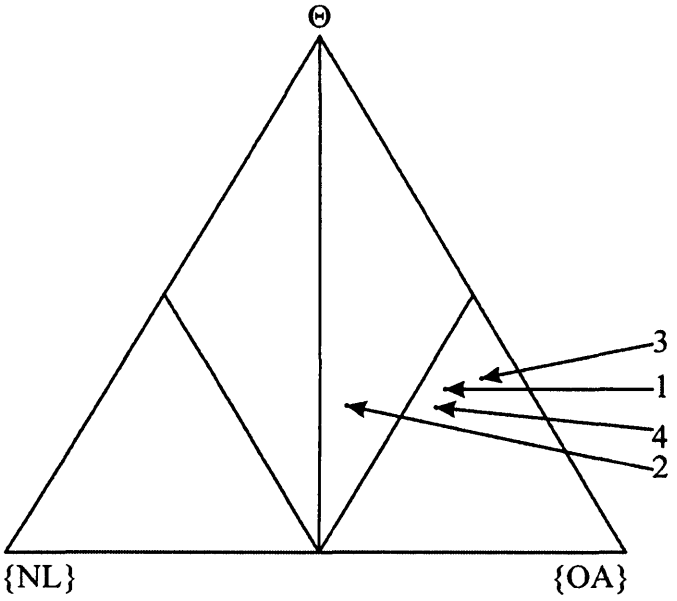


Figure 6.1 Simplex plot showing the simplex coordinate representations of the BOE_c for the four visits of P1.

The value of $m_c(\{OA\})$ at visit 3 has increased from visit 2. Indeed the value of $m_c(\{OA\})$ at visit 3 is greater than at visit 1. In the same way the value of $m_c(\{NL\})$ has decreased to a lower value than at visit 1. The simplex coordinate of visit 3 is situated within the dominant OA region of the simplex plot.

Finally at visit 4 the value of $m_c(\{OA\})$ has decreased and the value of $m_c(\{NL\})$ increased since visit 3. The simplex coordinate for visit 4 still lies within the dominant OA region of the simplex plot.

6.2.2 P1 BOA results

The observational gait analysis highlighted that at visit 1 P1 had a fixed flexion deformity of approximately 5°. P1 used a cane, and favoured the healthy knee when walking. Their walk was tentative with a very short stride length. P1 was unsteady when rising from a chair and their leg was stiff during passive flexion.

At visit 2 P1 still had a slight fixed flexion deformity but was able to walk much more confidently, evident by the increased speed of walking and the absence of a cane. It

was apparent that they were now putting more of their weight through the damaged knee. P1 was undergoing daily physiotherapy.

There were no significant changes in the observational gait analysis from visit 2 to visits 3 and 4.

6.2.3 P1 KOS results

The KOS scores for the four visits of P1 are recorded in Table 6.3. From visit 1 to visit 2 P1 recorded a great improvement in knee function as shown in the change in the KOS score from 25 to 63.75. A further improvement was recorded from visit 2 to visit 3 but this was less than that of the previous visit. At visit 4 P1 reported that their knee function was slightly worse than it was at visit 3.

Table 6.3 KOS scores, % for the four visits of P1.

Visit	KOS score, %
1	25
2	63.75
3	75
4	73.75

6.2.4 Discussion of results for P1

The increase in $m_c(\{NL\})$ and simultaneous decrease in $m_c(\{OA\})$ from visit 1 to visit 2 suggests that P1 experienced some relief from the symptoms associated with OA knee function and recovered some degree of NL knee function. This is in agreement with the BOA and the increase in the KOS scores, which correspond to an improvement in function.

From visit 2 to visit 3 the increase in $m_c(\{OA\})$ and decrease in $m_c(\{NL\})$ suggests that there was a "set-back" in recovery. Comparison with the BOE_c values at visit 3 with those at visit 1 suggests that the knee function was less NL at visit 3 than it was

at visit 1. This result does not correspond with the BOA, which suggests that there was no improvement in knee function, or the increase in the KOS scores, which suggest an improvement in knee function.

The increase in $m_c(\{NL\})$ and decrease in $m_c(\{OA\})$ from visit 3 to 4 imply that P1 experience some relief from the symptoms associated with OA knee function and recovered some level of NL knee function. As for the previous visit, these results contradict the conflicting findings of the KOS and BOA.

The DST results suggest that from preoperative to one year postoperative there has been a limited change in P1's knee function and recovery of NL knee function, and that in this case the prosthetic knee does not function in the same way as the NL knee. A Pearson's correlation test (SPSS 11, SPSS Inc.) revealed no significant correlation between the BOE_c values and the KOS score for P1 (correlation between $m_c(\{NL\})$ and KOS score, $r = 0.032$; correlation between $m_c(\{OA\})$ and KOS score , $r = 0.025$).

6.3 PATIENT 2 (P2)

P2 was followed before and at three stages after TKR surgery. The timing of these visits is given in Table 6.4.

Table 6.4 P2 visit summary.

Visit	Visit type
1	Preoperative
2	3 months postoperative
3	6 months postoperative
4	12 months postoperative

6.3.1 P2 DS classifier results

The BOE_c values for the four visits of P2 are recorded in Table 6.5. The simplex representations of the BOE_cs are depicted in Figure 6.2.

Table 6.5 BOE_c values for the four visits of P2.

BOE_c	Visit			
	1	2	3	4
$m_c(\{OA\})$	0.7383	0.6314	0.5502	0.4997
$m_c(\{NL\})$	0.0111	0.0964	0.1799	0.1582
$m_c(\Theta)$	0.2505	0.2721	0.2699	0.3421

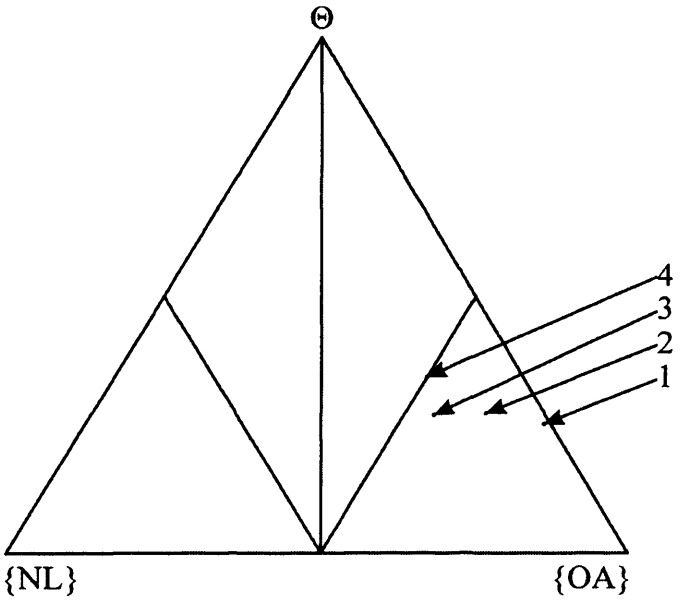


Figure 6.2 Simplex plot showing the simplex coordinate representations of the BOE_c for the four visits of P2.

At visit 1 P2 has a dominant OA classification since $m_c(\{OA\}) > [m_c(\{NL\}) + m_c(\Theta)]$. P2 has a very low $m_c(\{NL\})$ value. This is reflected in the positioning of the simplex coordinate in the simplex plot where it lies very near to the edge opposite the {NL} vertex.

At visit 2 the $m_c(\{OA\})$ value has decreased and the $m_c(\{NL\})$ value has increased. Despite this P2 continues to have a dominant OA classification as shown in the simplex plot.

At visit 3 there is a further decrease in $m_c(\{OA\})$ and an increase in $m_c(\{NL\})$. However, the simplex plot highlights that P2 still has a dominant OA classification.

Finally at visit 4 there is a decrease in both $m_c(\{OA\})$ and $m_c(\{NL\})$ from visit 3. However, the simplex coordinate associated with this visit has moved from the dominant to just inside the non-dominant OA region of the simplex plot.

6.3.2 P2 BOA results

At visit 1 P2 had a fixed flexion and varus deformity. Their gait was antalgic, stiff and jerky. These symptoms prevented them from walking more than a street length. They walked with a cane and placed less weight through the affected leg.

At visit 2 their leg was straighter in the frontal plane but they still had a fixed flexion deformity. P2 continued to favour the non-affected knee when walking as evident by the shorter stance phase on the side of the affected knee. However, it was evident that P2 was able to walk at a faster pace.

At visit 3 P2 was able to walk without a cane and at an increased pace. However, P2 still had a fixed flexion deformity. P2 continued to walk with a limp and place more weight through the un-affected knee.

There was no dramatic improvement from the observational gait analysis of visit 3 to visit 4.

6.3.3 P2 KOS results

The KOS scores for the four visits of P2 are recorded in Table 6.6. From visit 1 to visit 2 P2 reported a slight improvement in knee function. At visit 3 they reported a further improvement that was greater than that of the previous visit. At visit 4 P2 reported a worsening of knee function compared to visit 3.

Table 6.6 KOS scores for the four visits of P2.

Visit	KOS score, %
1	50
2	55
3	70
4	66.25

6.3.4 Discussion of results for P2

The increase in $m_c(\{NL\})$ and decrease in $m_c(\{OA\})$ from visit 1 to 2 suggests that P2 experienced some relief from the symptoms associated with OA knee function and recovered a level of NL knee function. This corresponds to the KOS and BOA results.

An increase in $m_c(\{NL\})$ and decrease in $m_c(\{OA\})$ again seen from visit 2 to visit 3 suggesting that P2 has experienced a further recovery of NL knee function and relief from the symptoms associated with OA knee function. This result is in agreement with both the BOA and KOS results.

The decrease in both $m_c(\{OA\})$ and $m_c(\{NL\})$ from visit 3 to visit 4 suggests that although there has been some loss of the characteristics associated with OA knee function this has not coincided with a recovery of NL knee function.

The overall change in the BOE_c values indicates that there has been some degree of recovery of NL knee function following TKR surgery. However, the non-dominant OA classification of P2 at visit 4 suggests that the prosthetic knee does not function in the same way as the NL knee. A Pearson correlation test revealed a significant correlation ($r = 0.966$, $p = 0.05$) between $m_c(\{NL\})$ and the KOS score.

6.4 PATIENT 3 (P3)

P3 was followed before and at three stages after TKR surgery. The timing of these visits is given in Table 6.7.

Table 6.7 P3 visit summary.

Visit	Visit type
1	Preoperative
2	2 months postoperative
3	6 months postoperative
4	15 months postoperative

6.4.1 P3 DS classifier results

The BOE_c values for the four visits of P3 are recorded in Table 6.8. The simplex representations of the BOE_cs are depicted in Figure 6.3.

Table 6.8 BOE_c values for the four visits of P3.

BOE_c	Visit			
	1	2	3	4
$m_c(\{OA\})$	0.8163	0.8323	0.7649	0.8213
$m_c(\{NL\})$	0.0067	0.0191	0.0486	0.0264
$m_c(\Theta)$	0.1770	0.1486	0.1865	0.1523

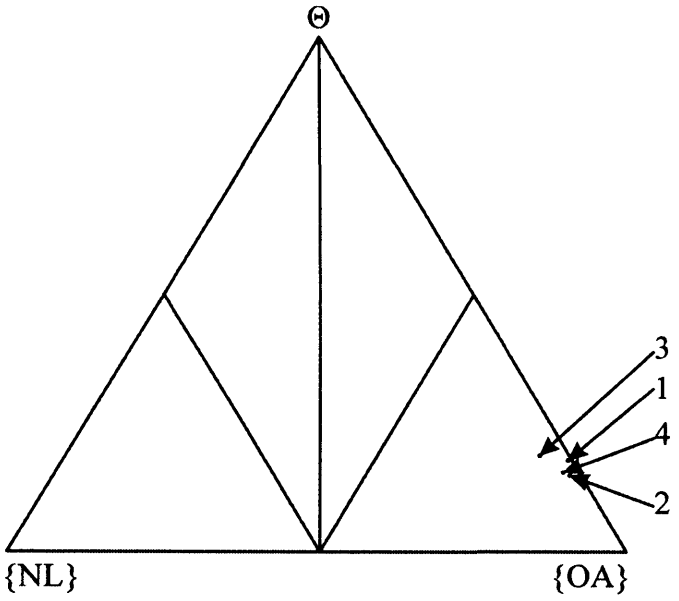


Figure 6.3 Simplex plot showing the simplex coordinate representations of the BOE_c for the four visits of P3.

At visit 1 P3 has a dominant OA classification since $m_c(\{OA\}) > [m_c(\{NL\}) + m_c(\Theta)]$. The $m_c(\{NL\})$ value is near zero which is reflected by the positioning of the simplex coordinate near to the edge opposite the {NL} vertex.

At visit 2 both the values of $m_c(\{OA\})$ and $m_c(\{NL\})$ have increased from visit 1. The simplex plot that the $m_c(\{OA\})$ value is still dominant.

At visit 3 there is an increase in the value of $m_c(\{NL\})$ and a decrease in $m_c(\{OA\})$.

At visit 4 $m_c(\{OA\})$ and $m_c(\{NL\})$ return to similar values as at visit 2. The simplex plot highlights that there has been little change in P3's knee function from their first visit.

6.4.2 P3 BOA results

At visit 1 P3 had a fixed flexion deformity of approximately 15° and was only able to achieve 90° in passive flexion. Additionally, P2 had a slight varus deformity. Their gait was antalgic and very slow. P2 was overweight and appeared to be very inactive which led the surgeon conducting the BOA to predict that the patient would be slow to recover from TKR surgery. P2 used their hands to aid in rising from a chair.

At visit 2 P3's leg was straighter in both the sagittal and frontal planes. P2 was now able to reach 110° in passive flexion. A greater range of motion was evident and their gait appeared to be less antalgic. P2 was now able to rise from a chair without using their hands.

At visit 3, P3 displayed a more cautious gait pattern than at visit 2.

There was no specific difference observed at visit 4 from visit 3.

6.4.3 P3 KOS results

The KOS scores for the four visits of P3 are recorded in Table 6.9. From visit 1 to visit 2 P3 reported a vast improvement in knee function as seen in the increase of over 50 points in the KOS results. This is followed by a setback in recovery as the KOS score decreases from visit 2 to visit 3. Finally, P3 reports an improvement in knee function from visit 3 to visit 4.

Table 6.9 KOS scores for the four visits of P3.

Visit	KOS score, %
1	31.25
2	82.5
3	58.75
4	68.75

6.4.4 Discussion of results for P3

The increase in $m_c(\{NL\})$ from visit 1 to visit 2 suggests that P4 has recovered some level of NL knee function. However, this increase in $m_c(\{NL\})$ coincided with a slight increase in $m_c(\{OA\})$, which suggests that P4 is displaying more signs of OA knee function than at visit 1. This increase in $m_c(\{NL\})$ corresponds to the BOA and KOS results which suggest an improvement in knee function.

From visit 2 to visit 3 the increase in $m_c(\{NL\})$ and decrease in $m_c(\{OA\})$ suggests that P3 experienced some relief from the symptoms associated with OA knee function and recovered a degree of NL knee function. However this does not correspond to the BOA or KOS results which both suggest that there was a setback in recovery.

The increase in $m_c(\{OA\})$ and decrease in $m_c(\{NL\})$ from visit 3 to visit 4 suggest that there was a setback in recovery. This result is not in agreement with the BOA result, which suggests that there was no change from visit 3 to 4, nor the KOS result, which suggests an improvement in knee function.

The DST results suggest that P3 experienced a limited recovery of NL knee function and thus the prosthetic knee does not function in the same way as a NL knee.

A Pearson's correlation test (SPSS 11, SPSS Inc.) revealed no significant correlation between the BOE_c values and the KOS score for P3 (correlation between $m_c(\{NL\})$ and KOS score, $r = 0.315$; correlation between $m_c(\{OA\})$ and KOS score , $r = 0.257$).

6.5 PATIENT 4 (P4)

P4 was followed before and at three stages after TKR surgery. The timing of these visits is given in Table 6.10.

Table 6.10 P4 visit summary.

Visit	Visit type
1	Preoperative
2	3 months postoperative
3	6 months postoperative
4	12 months postoperative

6.5.1 P4 DS classifier results

The BOE_c values for the four visits of P4 are recorded in Table 6.11. The simplex representations of the BOE_cs are depicted in Figure 6.4.

Table 6.11 BOE_c values for the four visits of P4.

BOE_c	Visit			
	1	2	3	4
$m_c(\{OA\})$	0.6938	0.6337	0.5721	0.5289
$m_c(\{NL\})$	0.0712	0.0955	0.1236	0.1732
$m_c(\Theta)$	0.2350	0.2708	0.3043	0.2979

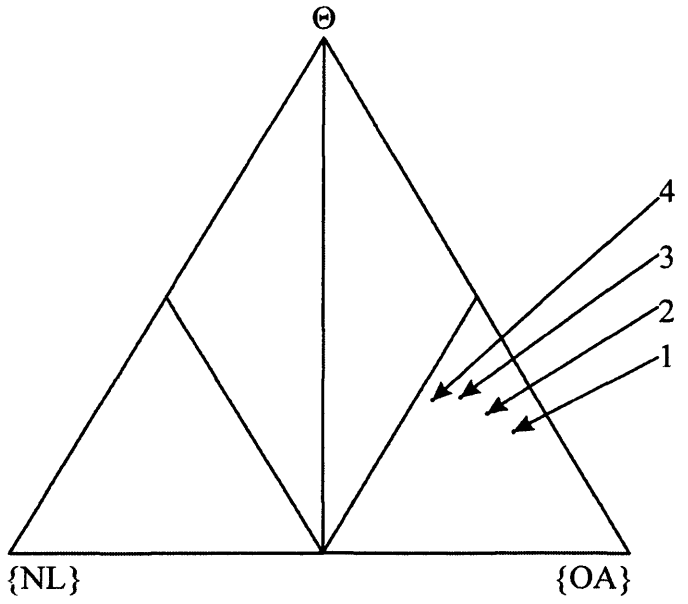


Figure 6.4 Simplex plot showing the simplex coordinate representations of the BOE_c for the four visits of P4.

At visit 1 P4 has a dominant OA classification since $m_c(\{OA\}) > [m_c(\{NL\}) + m_c(\Theta)]$.

At visit 2 P4 continues to have a dominant OA classification. However $m_c(\{OA\})$ has decreased and $m_c(\{NL\})$ increased since visit 1 suggesting that there has been a restoration of some NL function.

This pattern continues through to visits 3 and 4. This steady improvement in knee function is evident in the simplex plot where the simplex coordinate moves towards the edge opposite the {OA} vertex (associated with a decrease in $m_c(\{OA\})$) and away from the edge opposite the {NL} vertex (associated with an increase in $m_c(\{NL\})$).

6.5.2 P4 BOA results

At visit 1 P4 displayed an antalgic gait pattern. P4 had a fixed flexion deformity and walked with a stiff-leg gait. P4 was able to achieve 100° in passive flexion. P4 had a varus left knee and produced a varus thrust when walking. The patient was able to rise up out a chair easily.

At visit 2 P4 still displayed an antalgic stiff-leg gait. It was apparent that the patient was not putting full weight through their leg and the patient walked with a limp. Their knee was straighter in the frontal plane and their varus thrust had disappeared. P4 maintained their passive range of motion.

At visit 3 P4 continued to walk with a limp but took more weight through their leg than at visit 2. P4 walked at a faster pace than during their previous visits. The patient displayed signs of a painful right hip and back and the surgeon could not ascertain whether this affected P4's gait. There was no significant difference in their gait at visit 4 from visit 3.

6.5.3 P4 KOS results

The KOS scores for the four visits of P4 are recorded in Table 6.12. At each stage of recovery, P4 reported an improvement in knee function.

Table 6.12 KOS scores for the four visits of P4.

Visit	KOS score, %
1	28.75
2	40
3	46.25
4	56.25

6.5.4 Discussion of results for P4

At all stages of recovery there is an increase in $m_c(\{NL\})$ and a decrease in $m_c(\{OA\})$ suggesting that P4 experienced some relief from the symptoms associated with OA knee function and some recovery of NL knee function. These results are in agreement with the KOS results, which also suggest an increasing recovery of NL knee function. Pearson correlation tests identified strong correlations between $m_c(\{OA\})$ and the KOS score ($r = -0.990$, $p = 0.01$) and between $m_c(\{NL\})$ and the KOS score ($r = 0.979$, $p = 0.05$). Both the DST and KOS results are in agreement with the BOA results for visits 1 to 3 but differ for visit 4. Despite the increase in $m_c(\{NL\})$ through all stages of recovery P4 still had a dominant OA classification at visit 4 suggesting that the prosthetic knee does not function in the same way as the NL knee.

6.6 PATIENT 5 (P5)

P5 was followed before and at three stages after TKR surgery. The timing of these visits is given in Table 6.13.

Table 6.13 P5 visit summary.

Visit	Visit type
1	Preoperative
2	7 weeks postoperative
3	7 months postoperative
4	15 months postoperative

6.6.1 P5 DS classifier results

The BOE_c for the four visits of P5 are recorded in Table 6.14. The corresponding simplex coordinates are depicted in Figure 6.5.

Table 6.14 BOE_c values for the four visits of P5.

BOE_c	Visit			
	1	2	3	4
$m_c(\{OA\})$	0.4980	0.3241	0.2369	0.2206
$m_c(\{NL\})$	0.2200	0.2767	0.3538	0.3554
$m_c(\Theta)$	0.2820	0.3992	0.4093	0.4240

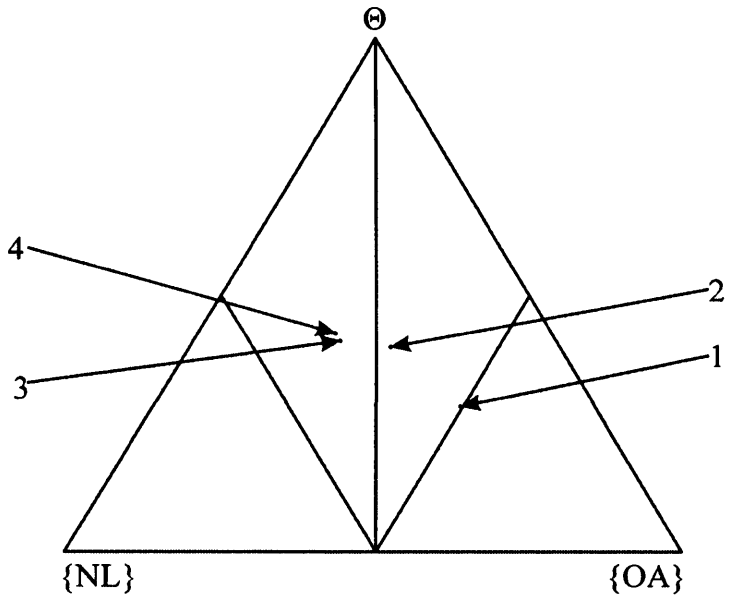


Figure 6.5 Simplex plot showing the simplex coordinate representations of the BOE_c for the four visits of P5.

At visit 1 P5 has a non-dominant OA classification. Although $m_c(\{OA\}) > m_c(\{NL\})$ and $m_c(\{OA\}) > m_c(\Theta)$, $m_c(\{OA\}) < [m_c(\{NL\}) + m_c(\Theta)]$. This is reflected in the positioning of the simplex coordinate for visit 1 within the non-dominant region of the simplex plot.

At visit 2 there has been an increase in $m_c(\{NL\})$ and a decrease in $m_c(\{OA\})$.

At visit 3 there has been a further increase in $m_c(\{NL\})$ and a decrease in $m_c(\{OA\})$ such that P5 now has a non-dominant NL classification. This change is evident from the simplex plot where the simplex coordinate lies in the non-dominant NL region.

Finally at visit 4 the values of $m_c(\{OA\})$ and $m_c(\{NL\})$ are very similar to those at visit 3 suggesting that there has been little change in knee function between the two visits.

6.6.2 P5 BOA results

P5 walked with an antalgic gait during visit 1. P2 was able to achieve 95° in passive flexion. The patient was confident when rising from a chair.

At visit 2 P5 walked with a slightly flexed knee during stance phase. P5's leg was stiff and favoured the unaffected knee during gait. P5 was able to achieve 105° in passive flexion.

At visit 3 P5 walked with a slight limp.

Finally, at visit 4 P5's leg appeared to be less stiff and the patient had an increased range of movement.

6.6.3 P5 KOS results

The KOS scores for the four visits of P5 are recorded in Table 6.15. P5 did not complete a KOS at visit 1. From visit 2 to visit 3 P5 reported no change in knee function, but reported an improvement from visit 3 to visit 4.

Table 6.15 KOS scores for the four visits of P5.

Visit	KOS score, %
1	-
2	56.25
3	56.25
4	73.75

6.6.4 Discussion of results for P5

At all stages of recovery there is an increase in $m_c(\{NL\})$ and decrease in $m_c(\{OA\})$ which suggests that P5 increasingly recovered some level of NL knee function. These findings are in agreement with the BOA results, which suggest a steady improvement in knee function. However, from the DS results, the level of uncertainty in the classification increased at each visit. A Pearson's correlation test (SPSS 11, SPSS Inc.) revealed no significant correlation between $m_c(\{OA\})$ and the KOS score (r = −0.621) or $m_c(\{NL\})$ and the KOS score (r = −0.817). The non-dominant NL classification at visit 4 suggests that P5 has recovered a greater degree of NL knee

function than any of the other patients. However, even in this case the prosthetic knee does not function in the same way as a NL knee.

6.7 PATIENT 6 (P6)

P6 was followed before and at three stages after TKR surgery. The timing of these visits is given in Table 6.16.

Table 6.16 P6 visit summary.

Visit	Visit type
1	Preoperative
2	3 months postoperative
3	6 months postoperative
4	12 months postoperative

6.7.1 P6 DS classifier results

The BOE_c values for the four visits of P6 are recorded in Table 6.17 and represented by their simplex coordinates in Figure 6.6.

Table 6.17 BOE_c values for the four visits of P6.

BOE_c	Visit			
	1	2	3	4
$m_c(\{OA\})$	0.7514	0.6753	0.4393	0.5920
$m_c(\{NL\})$	0.0561	0.0140	0.2906	0.1014
$m_c(\Theta)$	0.1925	0.3107	0.2701	0.3067

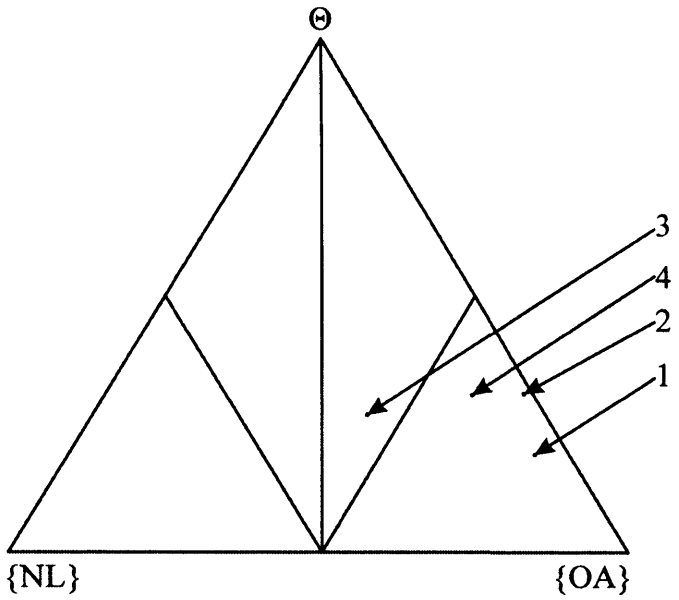


Figure 6.6 Simplex plot showing the simplex coordinate representations of the BOE_c for the four visits of P6.

At visit 1 P6 has a dominant OA classification with $m_c(\{OA\}) > [m_c(\{NL\}) + m_c(\Theta)]$. The simplex coordinate for visit 1 lies within the dominant OA classification region of the simplex plot.

At visit 2 there has been a decrease in $m_c(\{OA\})$ from visit 1; however $m_c(\{NL\})$ has also decreased. The level of uncertainty associated with this visit, $m_c(\Theta)$, has simultaneously increased.

At visit 3 $m_c(\{OA\})$ has decreased such that $m_c(\{OA\}) < [m_c(\{NL\}) + m_c(\Theta)]$. The value of $m_c(\{NL\})$ has also increased. This change is evident in the simplex plot where the simplex coordinate has moved to the non-dominant OA region.

Finally at visit 4 $m_c(\{OA\})$ has increased and $m_c(\{NL\})$ decreased.

6.7.2 P6 BOA results

At visit 1 P6 favoured the un-affected knee whilst walking evident in the shortened stance phase of the affected side. The patient appeared to walk with a stiff leg. However, it must be noted that the patient did not require the use of a cane.

At visit 2 the patient displayed a more stable and confident gait pattern. This was evident in an increased stride length. Despite the presence of swelling in the knee the patient was able to flex the knee more easily and the leg was less stiff than at visit 1.

At visit 3 and visit 4 P6 continued to show further signs of recovery compared to visit 2, particularly in the sagittal plane range of motion and the pace of walking.

6.7.3 P6 KOS results

The KOS scores for the four visits of P6 are recorded in Table 6.18. At each stage of recovery, P6 reported an improvement in knee function.

Table 6.18 KOS scores for the four visits of P6.

Visit	KOS score, %
1	36.25
2	55
3	71.25
4	60

6.7.4 Discussion of results for P6

From visit 1 to visit 2 the decrease in $m_c(\{OA\})$ and increase in $m_c(\{NL\})$ suggest that although P6 experienced some relief from the characteristics of OA knee function, this was not accompanied by a recovery of NL knee function. This result is not in agreement with the BOA and KOS results, which suggest an improvement in function.

From visit 2 to visit 3, the increase in $m_c(\{NL\})$ and decrease in $m_c(\{OA\})$ suggest that P6 experienced some relief from the symptoms associated with OA knee function and some recovery of NL knee function. This corresponds to the KOS and BOA results.

The increase in $m_c(\{OA\})$ and decrease in $m_c(\{NL\})$ from visit 3 to visit 4 suggests that there has been a setback in recovery. This is in agreement with the improvement in the KOS score reported by the patient but in contradiction to the BOA results, which suggest an improvement in knee function. At visit 4 P6 still had a dominant OA classification suggesting that after a year's recovery the prosthetic knee did not function in the same way as the NL knee.

A Pearson's correlation test (SPSS 11, SPSS Inc.) revealed no significant correlation between $m_c(\{NL\})$ and the KOS score (r = 0.728) nor between $m_c(\{OA\})$ and the KOS score (r = –0.944).

6.8 PATIENT 7 (P7)

P7 was followed before and at four stages after TKR surgery. The timing of these visits is given in Table 6.19.

Table 6.19 P7 visit summary.

Visit	Visit type
1	Preoperative
2	4 months postoperative
3	8 months postoperative
4	12 months postoperative

6.8.1 P7 DS classifier results

The BOE_c values for the four visits of P7 are recorded in Table 6.20. The simplex representations of the BOE_cs are depicted in Figure 6.7.

Table 6.20 BOE_c values for the four visits of P7.

BOE_c	Visit			
	1	2	3	4
$m_c(\{OA\})$	0.7847	0.6658	0.6452	0.6715
$m_c(\{NL\})$	0.0046	0.0621	0.0971	0.0619
$m_c(\Theta)$	0.2107	0.2721	0.2577	0.2667

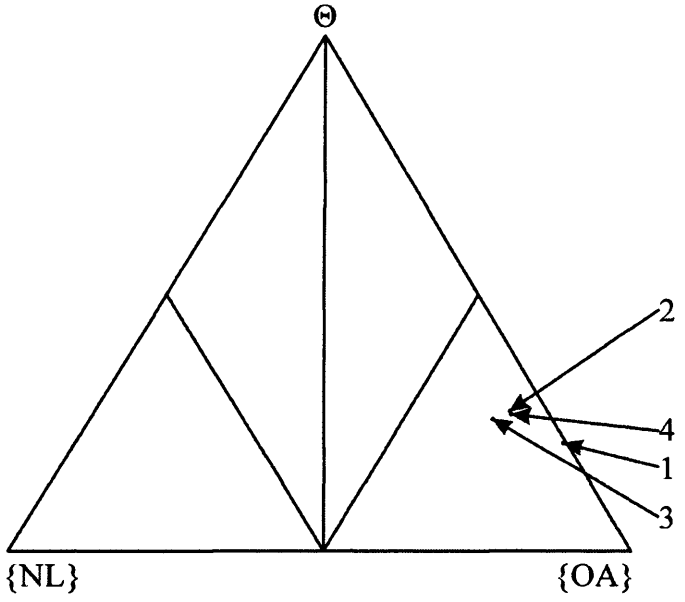


Figure 6.7 Simplex plot showing the simplex coordinate representations of the BOE_c for the four visits of P7.

At all stages of recovery P7 has a dominant OA classification since $m_c(\{OA\}) > [m_c(\{NL\}) + m_c(\Theta)]$. This is reflected in the positioning of the simplex coordinates within the dominant OA classification region.

At visit 1 $m_c(\{NL\})$ is near zero which is reflected in the positioning of the simplex coordinate near to the edge opposite the {NL} vertex.

From visit 1 to visit 2 there is an increase in $m_c(\{NL\})$ and a decrease in $m_c(\{OA\})$. This is again evident from visit 2 to visit 3.

Finally from visit 3 to visit 4 there is an increase in $m_c(\{OA\})$ and a decrease in $m_c(\{NL\})$ such that the BOE_c values are very close to those at visit 2.

6.8.2 P7 BOA results

At visit 1 P7 walked with a stiff-leg gait. P7 was able to achieve 95° in passive flexion. P7 relied on using their hands to rise up out of a chair. At each postoperative visit there was no overall change from visit 1.

6.8.3 P7 KOS results

The KOS scores for the four visits of P7 are recorded in Table 6.21. P7 did not complete a KOS at visit 1. From visit 2 to visit 3, and from visit 3 to visit 4, P7 reported an improvement in knee function.

Table 6.21 KOS scores for the four visits of P7.

Visit	KOS score, %
1	-
2	67.5
3	72.5
4	83.75

6.8.4 Discussion of results for P7

The increase in $m_c(\{NL\})$ and decrease in $m_c(\{OA\})$ from visit 1 to visit 2 suggests that P7 experienced a degree of improvement of NL knee function.

The increase in $m_c(\{NL\})$ and decrease in $m_c(\{OA\})$ from visit 2 to visit 3 suggests that P7 experienced further recovery of NL knee function. This result is in agreement with the KOS results.

The decrease in $m_c(\{NL\})$ and increase in $m_c(\{OA\})$ from visit 3 to 4 suggests that there has been setback in recovery. This is contrary to the KOS results, which suggest an improvement in function.

The overall change in the BOE_c values from visit 1 to visit 4 suggests that there has been a limited recovery of NL knee function. This is in agreement with the BOA results, which suggest limited improvement following TKR surgery.

A Pearson's correlation test (SPSS 11, SPSS Inc.) revealed no significant correlation between $m_c(\{NL\})$ and the KOS score ($r = -0.848$) nor between $m_c(\{OA\})$ and the KOS score ($r = -0.174$).

6.9 PATIENT 8 (P8)

P8 was followed before and at four stages after TKR surgery. The timing of these visits is given in Table 6.22.

Table 6.22 P8 visit summary.

Visit	Visit type
1	Preoperative
2	6 months postoperative
3	9 months postoperative
4	12 months postoperative

6.9.1 P8 DS classifier results

The BOE_c values for the four visits of P8 are recorded in Table 6.23. The simplex representations of the BOE_cs are depicted in Figure 6.8.

Table 6.23 BOE_c values for the four visits of P8.

BOE_c	Visit			
	1	2	3	4
$m_c(\{OA\})$	0.6947	0.7108	0.6987	0.6528
$m_c(\{NL\})$	0.0709	0.0215	0.0338	0.0112
$m_c(\Theta)$	0.2344	0.2678	0.2675	0.3360

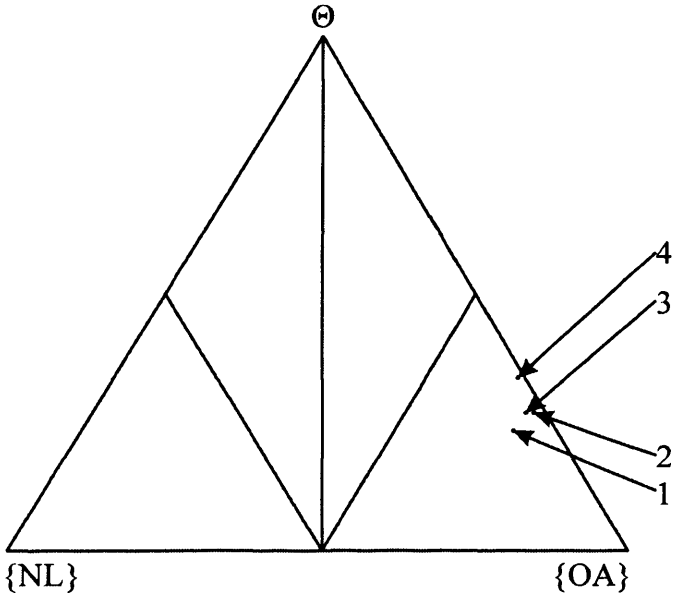


Figure 6.8 Simplex plot showing the simplex coordinate representations of the BOE_c for the four visits of P8.

At visit 1 P8 has a dominant OA classification since $m_c(\{OA\}) > [m_c(\{NL\}) + m_c(\Theta)]$. From visit 1 to visit 2 there is an increase in $m_c(\{OA\})$ and a decrease in $m_c(\{NL\})$. From visit 2 to visit 3 there is an increase in $m_c(\{NL\})$ and a decrease in $m_c(\{OA\})$. However $m_c(\{NL\})$ is lower than it was at visit 1. Finally from visit 3 to 4 there is a decrease in both $m_c(\{OA\})$ and $m_c(\{NL\})$. At each visit there is an increase in $m_c(\Theta)$.

6.9.2 P8 BOA results

At visit 1 P8 had a fixed flexion deformity of 10° and a slight varus deformity. P8 was able to achieve 90° in passive flexion. P8 walked with a stiff-leg gait and used their hands when rising from a chair. At visit 2 P8 had no fixed flexion deformity. Despite this, there was no real improvement from visit 1. Furthermore, there was no overall improvement from visit 2 to visit 3 or from visit 3 to visit 4.

6.9.3 P8 KOS results

The KOS scores for the four visits of P8 are recorded in Table 6.24. P8 did not complete a KOS for visit 2. From visit 1 to visit 3 and from visit 3 to visit 4, P8 reported an improvement in knee function.

Table 6.24 KOS scores for the four visits of P8.

Visit	KOS score, %
1	28.75
2	-
3	67.5
4	73.75

6.9.4 Discussion of results for P8

The increase in $m_c(\{OA\})$ and decrease in $m_c(\{NL\})$ from visit 1 to visit 2 suggests that there has been a reduction in NL knee function. This did not correspond to the BOA results, which suggested no change in knee function.

The increase in $m_c(\{NL\})$ from visit 2 to visit 3 suggests that P8 has experienced some relief from the symptoms associated with OA knee function and some recovery of NL knee function. However, comparison with the BOE_c values at visit 3 with those at visit 1 shows that there has been no improvement from their preoperative state.

The decrease in $m_c(\{OA\})$ and decrease in $m_c(\{NL\})$ from visit 3 to visit 4 suggest that although P8 experienced some relief from the symptoms associated with OA knee function there was no corresponding recovery of NL knee function.

The overall decrease in $m_c(\{OA\})$ and decrease in $m_c(\{NL\})$ from visit 1 to visit 4 suggests that P8 experienced some relief from the symptoms associated with OA knee function but also a loss of NL knee function. This suggests that the TKR surgery provides relief from the symptoms associated with OA knee function but that the prosthetic knee does not function in the same way as the NL knee.

A Pearson's correlation test (SPSS 11, SPSS Inc.) revealed no significant correlation between $m_c(\{NL\})$ and the KOS score ($r = -0.223$) nor between $m_c(\{OA\})$ and the KOS score ($r = 0.414$).

6.10 PATIENT 9 (P9)

P9 was followed before and at three stages after TKR surgery. The timing of these visits is given in Table 6.25.

Table 6.25 P9 visit summary.

Visit	Visit type
1	Preoperative
2	3 months postoperative
3	7.5 months postoperative
4	12 months postoperative

6.10.1 P9 DS classifier results

The BOE_c values for the four visits of P9 are recorded in Table 6.26. The simplex representations of the BOE_c are depicted in Figure 6.9.

Table 6.26 BOE_c values for the four visits of P9.

BOE_c	Visit			
	1	2	3	4
$m_c(\{OA\})$	0.3821	0.5820	0.1735	0.3450
$m_c(\{NL\})$	0.2086	0.1752	0.3933	0.2248
$m_c(\Theta)$	0.4093	0.2428	0.4332	0.4302

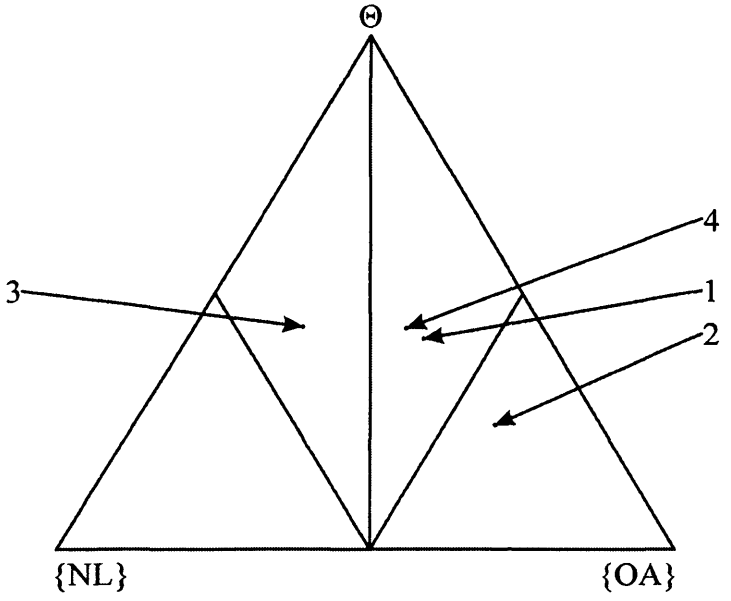


Figure 6.9 Simplex plot showing the simplex coordinate representations of the BOE_c for the four visits of P9.

At visit 1 $m_c(\{OA\}) > m_c(\{NL\})$ but $m_c(\{OA\}) < [m_c(\{NL\}) + m_c(\Theta)]$ and thus the simplex coordinate for visit 1 lies within the non-dominant region of the simplex plot.

At visit 2 $m_c(\{OA\})$ has increased and $m_c(\{NL\})$ decreased since visit 2. P2 now has a dominant OA classification as indicated by the positioning of the simplex coordinate within the dominant OA region of the simplex plot.

At visit 3 $m_c(\{OA\})$ has decreased to below its value at visit 1 and $m_c(\{NL\})$ has increased to above its value at visit 1. At this visit $m_c(\{NL\}) > m_c(\{OA\})$ and the simplex coordinate has moved to the non-dominant NL region of the simplex plot.

Finally, at visit 4 $m_c(\{OA\})$ has increased and $m_c(\{NL\})$ has decreased such that the simplex coordinate has moved back to the non-dominant NL region of the simplex plot. The simplex coordinate of visit 4 lies near to the simplex coordinate of visit 1.

6.10.2 P9 BOA results

At visit 1 P9 had a slight varus knee and produced a varus thrust at heel strike. The patient did not exhibit signs of pain during walking or rising from a chair. The patient appeared to be fit and healthy and the surgeon expected that P9 would cope well with the TKR.

At visit 2 P9 had a fixed flexion deformity of approximately 5°. P9 displayed an antalgic and cautious gait pattern. This was particularly evident in the shortened stance phase on the affected side. P9 had a greater range of motion and was able to achieve 120° in passive flexion. P9 used their hands to balance when rising from the chair and favoured their unaffected knee when standing.

At visit 3 P9 walked with a more confident gait pattern, returning back to the pattern evident during visit 1. P9's leg was straighter in the sagittal plane although they were still not able to achieve full range of motion.

At visit 4 there was a slight regression from visit 3. P9's leg was stiffer than at visit 3 and P9 still had a fixed flexion deformity and was not able to achieve full range of motion.

6.10.3 P9 KOS results

The KOS scores for the four visits of P9 are recorded in Table 6.27. From visit 1 to visit 3 P9 reported a worsening of knee function. From visit 2 to visit 3 P9 reported an

improvement in knee function. This was followed by a "set-back" in recovery from visit 3 to visit 4.

Table 6.27 KOS scores for the four visits of P9.

Visit	KOS score, %
1	68.75
2	66.25
3	80
4	70

6.10.4 Discussion of results for P9

The increase in $m_c(\{OA\})$ and decrease in $m_c(\{NL\})$ from visit 1 to visit 2 suggest that P9 experienced an increased function associated with OA. This corresponds to a decrease in the KOS results and the findings of the BOA.

The increase in $m_c(\{NL\})$ and decrease in $m_c(\{OA\})$ from visit 2 to visit 3 suggests that P9 experienced a degree of recovery of NL knee function. This result is in agreement with the KOS and BOA results.

Finally, from visit 3 to visit 4 the increase in $m_c(\{OA\})$ and decrease in $m_c(\{NL\})$ suggests that there has been a "set-back" in recovery. Comparison of the BOE_c values at visit 4 with those at visit 1 suggests that between visit 3 and 4 there has been a regression towards the preoperative state. This result is in agreement with the BOA and KOS results. A Pearson correlation test revealed a very strong relationship between $m_c(\{NL\})$ and the KOS score ($r = 0.999, p = 0.01$).

6.11 DISCUSSION OF OVERALL RESULTS

This study was conducted to examine the ability of the DS classifier to identify changes in knee function following TKR surgery. The results for nine patients have been presented and discussed. A set of requirements that a tool for the assessment of

the outcome of TKR surgery should fulfil, was given by Davies (2002) (see section 6.1). A discussion of the overall results from this study is now given in reference to these requirements.

6.11.1 Establishing the level of benefit achieved by surgery

Results showed that the BOE_c values enable the level of benefit achieved by surgery to be established as discussed in sections 6.2 to 6.10. Although the DS results suggest that most of the patients showed some degree of recovery of NL knee function, none of the patients at any stage of recovery gained a dominant NL classification. This suggests that none of the patients recovered complete NL knee function during level walking following TKR surgery. This is in agreement with the work of Benedetti *et al.* (2003), Andriacchi (1993), Myles *et al.* (2002), Fuchs *et al.* (2002) and Whittle and Jefferson (1989) who reported that TKR patients do not achieve NL knee function over time. Since working with numerous TKR patients during the knee clinical trial (see Chapter 2), this lack of restoration of NL function hardly seems surprising considering the alteration in joint structure, the deterioration of muscles and soft tissue and the overall condition of the patients.

During TKR surgery, the articulating surfaces of the tibia, femur and sometimes the patella are replaced with prosthetic components. Additionally, the anterior cruciate ligament (ACL) and in some cases the posterior cruciate ligament (PCL) are sacrificed. Consequently, the NL biomechanics of the knee are altered.

In constrained TKR designs, there is an increased congruity of the prosthetic surfaces, which increases the stability of the knee joint and restrains motion. Additionally, in all TKR designs the friction between the metal and plastic surfaces has an effect on the kinematics of the knee (Walker and Sathasivam, 2000) again resulting in increased knee stability.

The PCL restrains the anterior movement of the femur on the tibia. During flexion, the PCL displaces the femur in a posterior direction, which increases the lever arm of the quadriceps muscle. Removal of the PCL restrains the tibio-femoral rollback during flexion and consequently reduces the lever arm of the quadriceps. This

increases the demand of the quadriceps muscle. However, in patients who suffer from muscle atrophy, the quadriceps muscle may not be able to provide the additional force needed to extend the knee. Andriacchi (1993) reported that patients compensate for this deficiency by leaning the body forwards. This reduces the knee flexion moment and consequently reduces the demand on the quadriceps. Additionally the PCL takes one third of the shear force transmitted through the knee (Walker and Sathasivam, 2000). Removal of the PCL means that the shear forces on the surfaces of the prosthetic joint are increased.

During mid stance, when the knee is near full extension, the contraction of the quadriceps muscle produces an anterior pull of the patella ligament on the tibia. This is stabilised by the ACL. In ACL deficient knees, patients compensate for no ACL by reducing the demand on the quadriceps, which is seen in a reduction in the flexion-extension moment (Andriacchi, 1993). It has been reported that patients with uni-compartmental knees, which allow for the retention of the ACL, obtain a function that closely resembles that of a NL knee (Andriacchi, 1993).

Many patients with knee OA adopt compensatory gait mechanisms to reduce the pain experienced during walking. It seems apparent that, over time, these adaptations become habitual. This became particularly evident whilst working on the knee clinical trial when patients continued to walk with compensatory gait mechanisms after TKR surgery.

Several of the TKR patients that participated in the clinical trial suffered from co-morbidities namely angina, shortness of breath, and hip, spine and ankle pathologies. It is inevitable that these additional problems would affect the patient's ability to walk with a NL gait pattern.

The lack of restoration of NL function following TKR surgery has two main implications. Firstly there will be a resultant effect on other joints, e.g., the other knee, hip, back and ankle because these joints have to compensate for the limitations of the replaced knee joint. Secondly, the abnormal kinematics and consequent increase in the shear forces on the surfaces of the prosthetic joint leads to excessive wear of the prosthetic components and subsequent loosening of the prosthesis.

6.11.2 Comparison of subjects

Additionally, the BOE_c values make comparison of different subjects possible. The preoperative and final postoperative BOE_c values for the 9 TKR patients are recorded in Table 6.28 in which the subjects have been ranked in descending order of $m_c(\{NL\})$. Both preoperative and postoperative the patients have varying levels of NL knee function. Preoperative, P5 shows the greatest level of NL knee function and P7 the least. Postoperative, P5 shows the greatest level of NL knee function and P8 the least. In general, the patients with the greatest levels of NL knee function before TKR surgery exhibit the greatest levels of NL knee function after surgery as is expected.

Table 6.28 Pre and post-operative BOE_c values for the 9 patients (P1-P9). Patients ranked in descending order of pre-operative $m_c(\{NL\})$.

Patient	$m_c(\{NL\})$		$m_c(\{OA\})$		$m_c(\Theta)$	
	Pre-op	**Post-op**	**Pre-op**	**Post-op**	**Pre-op**	**Post-op**
P5	0.2200	0.3554	0.4980	0.2206	0.2820	0.4240
P9	0.2086	0.2248	0.3821	0.3450	0.4093	0.4302
P1	0.1366	0.1718	0.5469	0.5489	0.3164	0.2793
P4	0.0712	0.1732	0.6938	0.5289	0.2350	0.2979
P8	0.0709	0.0112	0.6947	0.6528	0.2344	0.3360
P6	0.0561	0.1014	0.7514	0.5920	0.1925	0.3067
P2	0.0111	0.1582	0.7383	0.4997	0.2506	0.3421
P3	0.0067	0.0264	0.8163	0.8213	0.1770	0.1523
P7	0.0046	0.0619	0.7847	0.6715	0.2107	0.2667

6.11.3 Relating the outcome to clinical results

The results of the DS classifier were compared to the results of a KOS as discussed in sections 6.2 to 6.10. Overall, there was no significant correlation between the BOE_c values and the KOS scores. In the only other study that attempted to relate gait

analysis parameters to patient-related scoring systems, Fuchs *et al.* (2002) found no correlation between the two measures of outcome. None of the three methods (DST, KOS and BOA) were consistent with each other in providing similar outcomes. Several reasons may exist for this disparity. Firstly, the DST results may be affected by co-morbidity since some of the input variables used are not specifically knee function measures, e.g., BMI and cadence. In contrast, the BOA and KOS concentrate specifically on the knee function. Secondly, each of the three methods provides a different perspective on the assessment of knee function. The DST method assesses knee function during level walking. In contrast, the KOS score measures the function of the knee during different daily activities and considers a set of clinical parameters e.g. buckling, instability, pain etc. In addition to assessing knee function during level walking the BOA analysis took into account the static alignment of the knee, passive range of motion and the knee function whilst rising from a chair. It is also worth noting that the surgeon appeared to consider what was achievable by the individual patient and measured improvement on these terms rather than by comparison with the NL population. The results raise the question whether the outcome of TKR surgery should be measured in terms of patient wellbeing or technical success?

6.11.4 Using important measurable characteristics of the knee

The variables used as inputs to the DS classifier were measured during the knee function clinical trial as described in Chapter 2. These variables include anthropometrical measurements, temporal-distance parameters and kinetic and kinematic temporal waveforms. These variables are ones that have been cited in the literature as clinically relevant as discussed in Chapter 7 (section 7.6). However, when comparing the classifier outcomes with the subjective clinical opinions and KOS scores there are differences in the input variables used. Further work is required to study this aspect if the DS classifier is to be used in comparison with clinical and quality-of-life scores.

6.11.5 Simplicity

The DS method is simple and logical and the progression from taking clinically relevant measurements to making a decision using the simplex plot can be clearly

followed. This has become evident when communicating the method and results across a wide spectrum of disciplines (Beynon *et al.*, 2002; Jones *et al.*, 2003a, Jones *et al.*, 2003b, Jones *et al.*, 2004). The method has been well received by engineers, mathematicians, physiotherapists and orthopaedic surgeons alike. However, the method requires further thought, development and validation before it can be implemented.

6.11.6 Comparison between different surgical techniques or implants

It is anticipated that the tool can be used to compare outcomes from different surgical techniques or implants, although this prospect is beyond the scope of the current study. It is proposed to use the method to study the differences between rotating platform and fixed bearing knee implants taking into account the comments from this chapter.

This chapter has investigated the use of the DS classifier as a tool for assessing the outcome of TKR surgery. The results for a set of nine patients have been presented and discussed. Conclusions taken from and further work based on this study are presented in Chapter 8.

CHAPTER 7

OTHER CLASSIFICATION METHODS

In Chapter 1 an introduction to the different methods used in gait classification studies was given. This chapter presents a comparison of the Dempster-Shafer (DS) classifier with two of these classification methods: the Artificial Neural Network (ANN) and Linear Discriminant Analysis (LDA). The first classifier, the ANN is chosen as a basis for comparison since, like the DS classifier, it is a non-linear method. Additionally the ANN has been used extensively in classification studies in recent years (see section 1.2.3). In contrast to the first classifier, the second classifier, LDA is a linear method. It is a method that is well established and has been used as a benchmark in other comparative studies (Gioftsos and Grieve, 1995; Wu and Su, 2000). The comparison is centred on two main aspects: **prediction** and **interpretability** (Breiman, 2001). In this context, prediction relates to the ability of the classifier to assign new subjects to the correct class and is vital to ensure confidence in its use. This can be measured in terms of the out-of-sample classification accuracy. Interpretability relates to the ease with which the user can extract meaningful information as to the relationship between the output of the classifier and the input data. Following Jones *et al.* (2004) this is measured using a ranking of variables.

The chapter begins with an introduction to the ANN and its application to the classification of osteoarthritic (OA) and normal (NL) knee function in section 7.1. Subsequently, the classification results are presented in terms of accuracy and interpretability in section 7.2. The same pattern is followed in sections 7.3 and 7.4 for the LDA classifier. Section 7.5 provides a summary of the DS classifier accuracy and interpretability results. Finally, a discussion based on a comparison of the results from the three classifiers is given in section 7.6.

7.1 THE ARTIFICIAL NEURAL NETWORK (ANN)

An overview of the application of ANNs to gait classification was given in Chapter 1. A summary of the ANNs used in these studies is given in Table 7.1.

Although many types of ANNs exist, it can be seen from this summary that the most commonly used ANN in gait classification is the feed-forward ANN trained with the back-propagation algorithm. Consequently, this type of ANN is used as the basis for this comparative study. Any particular ANN is specified by its net topology and training rules. These will now be described for the feed-forward ANN trained with the back-propagation algorithm.

7.1.1 Network topology

The ANN comprises processing elements, neurons, which mimic the function of the biological neurons in the brain. Within the ANN, these neurons are grouped together in layers. Generally, the ANN comprises an input layer, in which input neurons are connected to the input data; one or more hidden layers, which are the processing layers; and an output layer in which output neurons transmit the final output to the user. Although Barton and Lees (1995, 1997) utilised ANNs with two hidden layers, it has been reported in the literature that one hidden layer is sufficient for an ANN to learn any continuous relationship between the inputs and outputs (Bishop, 1995; Cybenko, 1989; Chau, 2000b). Consequently, this comparative study is restricted to one hidden layer.

The ANN with one hidden layer is depicted in Figure 7.1. The input layer consisting of input neurons, I_i (i = 1: n_i) is connected to the input data, v_i (i = 1: n_i) and has the same number of input neurons as there are input variables. The hidden layer consisting of hidden neurons, H_j (j = 1: n_j) is connected to the input layer by a set of connecting weights, W_{ji}. The output layer consists of output neurons, O_k (k = 1: n_k) and is connected to the hidden layer via connecting weights, U_{kj}. For a given layer each neuron is connected to all neurons in the preceding and subsequent layers. In the feed-forward ANN, neurons in one layer pass their output to the neurons in the subsequent layer but not to the preceding layer. Thus, information passes through the network in a forward-direction only.

Table 7.1 Summary of ANN studies used by other researchers in application to gait classification

Study	Network	Number of hidden layers	Number of neurons in hidden layer	Learning rate	Momentum term	Stopping criterion	Number of runs
Barton and Lees, 1995	Feedforward	2	10 (1^{st}), 5 (2^{nd})	0.6	0.8	MSE = 0.07	1
Barton and Lees, 1997	Feedforward	2	5 (1^{st}) ,4 (2^{nd})	–	–	–	4
Gioftsos and Grieve, 1995	Recurrent	1	9	0.02	0.9	Root mean square error	2
Holzreiter and Köhle, 1993	Feedforward	1	–	0.2	–	200,000 iterations	20
Lafuente *et al.*, 1998	Feedforward	1	3, 6, 9, 12, 15, 20, 25	0.4 – 0.6	–	75 epochs	6
Su and Wu, 2000	Feedforward	1	5, 10, 15, 20, 25	–	–	MSE = 0.01	10
Wu and Su, 2000	Feedforward	1	5, 10, 15, 20, 25	–	–	–	several

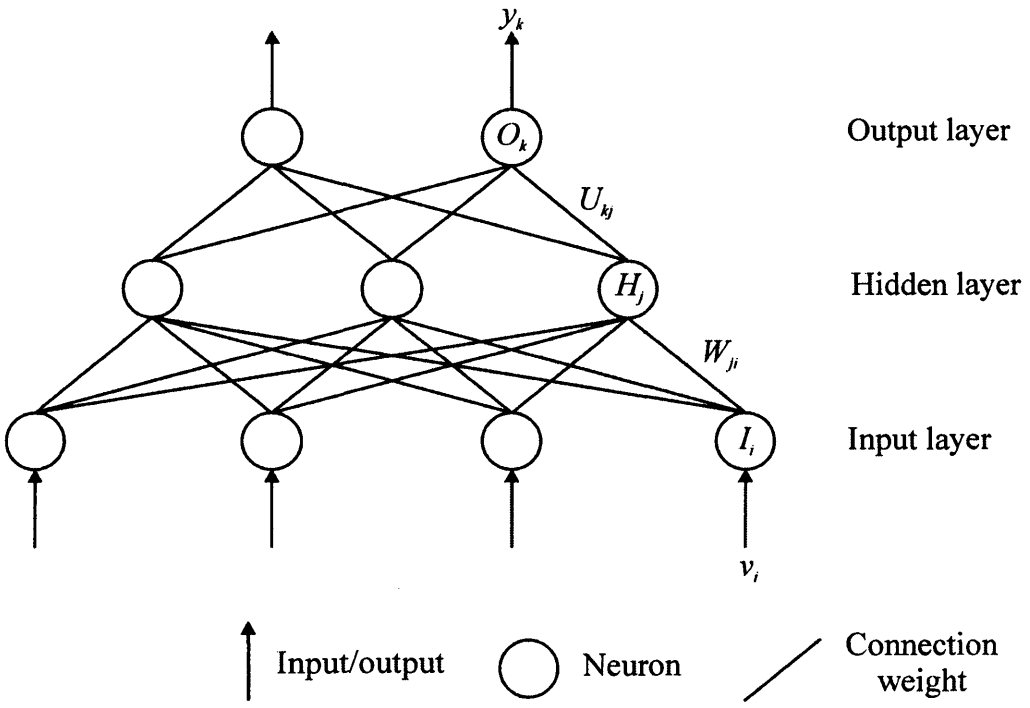


Figure 7.1 The ANN is arranged in layers: an input layer consisting of input neurons I_i; the hidden layer consisting of hidden neurons H_j; and the output layer consisting of output neurons O_k. The layers are connected by connecting weights: the weights connecting input neurons to hidden neurons are labelled W_{ji} and the weights connecting hidden neurons to the output neurons are labelled U_{kj}. The input neurons are connected to the input data v_i and the output neurons transmit the network outputs y_k.

Neurons in the input layer pass the input data to the hidden layer without performing any transformations (Figure 7.2a). A single neuron in the hidden layer receives a set of inputs v_i ($i = 1: n_i$) via the weighted connections, W_{ji} as shown in Figure 7.2b. The hidden neuron computes a weighted sum of all the inputs, a_j:

$$a_j = \sum_i^{n_i} W_{ji} v_i \tag{7.1}$$

The weighted sum is subsequently passed through an activation function, $f(.)$ which produces the neuron output, z_j. Following previous studies (Holzreiter and Köhle, 1993; Barton and Lees, 1995; Su and Wu, 2000; Wu and Su, 2000) a log-sigmoid activation function was utilised. This activation function transforms a_j into a value in the interval [0, 1]:

$$z_j = \frac{1}{1+e^{-a_j}} \tag{7.2}$$

In a similar way (Figure 7.2c), a single output neuron transforms its inputs z_j ($j = 1: n_j$) into an output y_k using

$$y_k = \frac{1}{1+e^{-b_k}} \tag{7.3a}$$

where

$$b_k = \sum_j^{n_j} U_{kj} z_j \tag{7.3b}$$

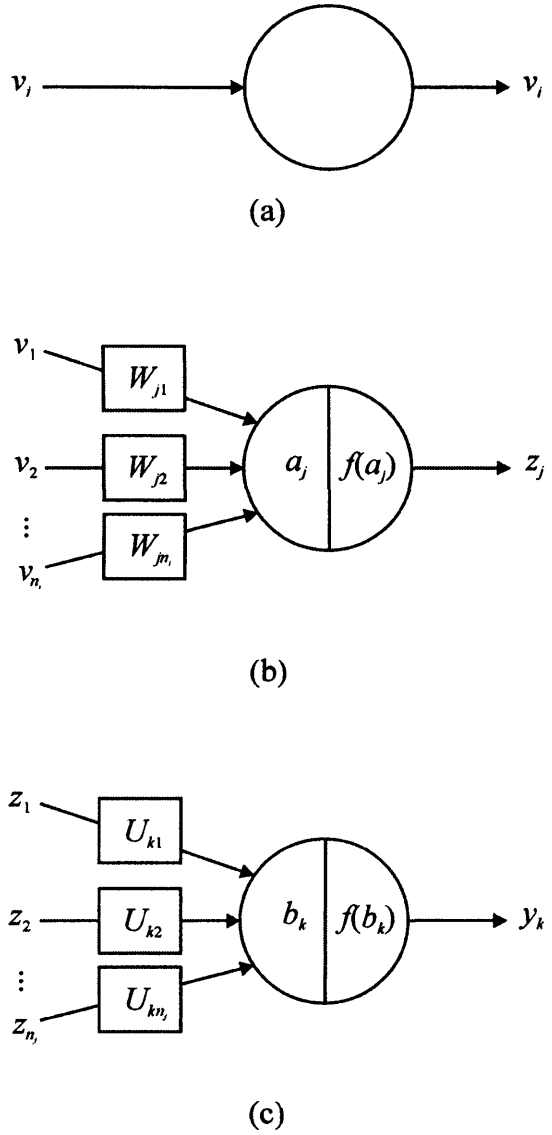


Figure 7.2 a) The input neuron passes an input variable v_i to the hidden layer; b) The hidden neuron receives a set of inputs, v_i (i = 1: n_i) via weighted connections, W_{ji}. The sum of the weighted inputs, a_j is subsequently passed through an activation function, $f(a_j)$ which produces the neuron output z_j; c) A similar process is undertaken by the output neuron on the inputs z_j (j = 1: n_j) to produce a sum of weighted inputs b_k and finally a neuron output y_k.

The output neurons, y_k ($k = 1: n_k$) are used in the classification of subjects. The way in which subjects are assigned to their respective classes is dependent on the number of output neurons. For an ANN with a single output neuron, subjects are assigned to a specific class based on the value of the output from that single neuron (see Holzreiter and Köhle, 1993). A consequence of using a sigmoid transfer function is that the output value is in the range [0,1]. In application to the current work of classifying NL and OA knee function, output values in the upper half of this range correspond to a subject with OA knee function, whilst those nearer to 0 correspond to a subject with NL knee function. For an ANN with more than one output neuron, subjects are assigned to a specific class based on the values of all the outputs (for an example of three outputs the reader is directed to Barton and Lees, 1995). For two-group classification of OA and NL knee function a maximum of two output neurons are used. The first output neuron corresponds to OA knee function whilst the second output neuron corresponds to NL knee function. If the value of the first output neuron is greater than that of the second output neuron, a subject is assigned to the OA class and vice-versa.

Since random values are initially assigned to the connecting weights, it is unlikely that the network will produce a desired output for the given inputs and so the values of the connection weights must be adjusted. This is achieved through a process called training. This process will now be described in detail.

7.1.2 Training Rules

There are two different ways in which the ANN can be trained: using supervised or unsupervised training. In this comparative study supervised training is used (see all studies in Table 7.1). During supervised training, examples of inputs and their corresponding outputs (target outputs) are presented to the ANN. By iteratively adjusting the connection weights, the ANN learns to infer the relationship between the inputs and outputs. Once the network has been trained, it can accept previously unseen data and attempt to predict an associated output, simply by evaluating the internal functions. The weights can be adjusted after a single pattern has been presented to the network (incremental training) or after the entire training set has been

presented (batch training). In this study, supervised batch training is achieved with the backpropagation algorithm.

Using the backpropagation algorithm the weights are adjusted according to the derivative of some error function, E. Following the work of Barton and Lees (1995) and Su and Wu (2000) the mean squared error (MSE) function was utilised:

$$E = \frac{1}{n_k}\sum_{k=1}^{n_k}(y_k - t_k)^2 \tag{7.4}$$

This error is a measure of the difference between the target output, t_k and the actual neuron output, y_k for the output neuron, O_k.

For a full exposition of the backpropagation algorithm the reader is directed to Bishop (1995). A brief summary is now given (Bishop, 1995).

1. Calculate the errors δ_k for each output unit using:

$$\delta_k = y_k(1 - y_k)(t_k - y_k) \tag{7.5}$$

2. Backpropagate these errors through the network to evaluate the errors δ_j for each hidden neuron using:

$$\delta_j = z_j(1 - z_j)\sum_{k=1}^{n_k} W_{kj}\delta_k \tag{7.6}$$

3. Evaluate the derivatives $\frac{\partial E}{\partial W_{ji}}$ and $\frac{\partial E}{\partial U_{kj}}$ using:

$$\frac{\partial E}{\partial W_{ji}} = \delta_j v_i \tag{7.7a}$$

$$\frac{\partial E}{\partial U_{kj}} = \delta_k z_j \tag{7.7b}$$

4. Update the weights after all the patterns in the training set have been presented to the network, according to the following:

$$\Delta W_{ji(t)} = -\eta \sum_{n=1}^{n_p} \left(\delta_j v_i\right)_n + \mu \Delta W_{ji(t-1)} \tag{7.8a}$$

$$\Delta U_{kj(t)} = -\eta \sum_{n=1}^{n_p} \left(\delta_k z_j\right)_n + \mu \Delta U_{kj(t-1)} \tag{7.8b}$$

where n_p is the number of patterns in the training set, η the learning rate and μ the momentum constant. The learning rate controls the amount by which the weights are updated. If the learning rate is too small then the algorithm will take a long time to converge to the global minimum, whilst if the value is too large the algorithm may not converge at all and get stuck in local minima. Introducing a momentum term can speed up convergence to the global minimum and avoid the pitfalls of convergence to local minima.

The neural network is trained until some stopping criterion is met (Bishop, 1995). Following the work of Barton and Lees (1995), this comparative study uses a pre-set MSE value combined with a limit on CPU time of 1 hour. Once the network has been trained it must be evaluated. This is achieved using a leave-one-out cross validation (LOOCV) approach as described in Chapter 3 (section 3.6).

7.1.3 Contribution of input variables

The ANN is generally perceived to be a "black box" approach to classification, that is to say, its "unimaginably complex inner workings somehow magically transform inputs into predicted outputs" (Garson, 1991, pp.47). For ANN with hidden layers the relationship between the inputs and the predicted outputs is difficult to realise mathematically because of their internal structure. As a result, it is difficult to understand the contribution that each input variable makes to the final classification of

subjects. Different interpretation methods have been suggested to capture the relationship between the input variables and the output units. One of these methods, suggested by Yoon *et al.* (1993) is used here to assess the contribution that each variable makes to the overall classification. This contribution is measured through examination of the internal weights of the ANN. The relative strength, RS_{ki} of the relationship between an input variable v_i and an output unit O_k, is given by

$$RS_{ki} = \frac{\sum_{j=1}^{n_j} (W_{ji} \times U_{kj})}{\sum_{i=1}^{n_i} \left| \sum_{j=1}^{n_j} (W_{ji} \times U_{kj}) \right|} \tag{7.9}$$

7.2 RESULTS OF THE ANN CLASSIFER

7.2.1 ANN Accuracy Results

To determine the most appropriate network topology and training strategy a series of tests were carried out to evaluate the discrimination and prediction capabilities of a number of different ANN. A summary of these tests is given in Table 7.2. In each set of tests one parameter was altered whilst the rest were kept constant in order to determine the effect of the given parameter on performance. The majority of the parameters used are to be found in the literature (see Table 7.1 and Duda *et al.*, 2001). The default parameters used were 5 hidden neurons, a learning rate of 0.1, a momentum constant of 0.9, a training goal (MSE) of 0.01 and a single output neuron. Since the connection weights are initially randomised, ten training runs were performed for each network configuration. For each run of the algorithm 42 ANN were created (one for each left out person). The results corresponding to a single run of the algorithm are therefore an average of these 42 ANN. The ANNs were programmed using Matlab 5.3 software (The Mathworks Inc.) and run on a desktop computer.

Table 7.2 Description of tests carried out to determine best network topology and training strategy

Test	Parameter varied	Part	Value of parameter
1	Training goal	a	0.07
		b	0.05
		c	0.01
		d	0.005
		e	0.001
2	Learning Rate	a	0.02
		b	0.1
		c	0.6
3	Momentum Constant	a	0.9
		b	0.8
4	Number of hidden neurons	a	5
		b	10
		c	15
		d	20
5	Number of output neurons	a	1
		b	2

7.2.1.1 Test 1 results – Effect of changing the training goal on classification accuracy

The results of test 1 to establish the effect of the training goal parameter on the in- and out-of-sample accuracy are recorded in Appendix C (Tables C.1 to C.5 respectively). The average and standard deviation of the ten training runs for the in- and out-of-sample accuracies are summarised in Table 7.3. A measure of the discrepancy

between the in- and out-of-sample performance, named the prediction bias (Lafuente *et al.*, 1998), is given in Figure 7.3.

Table 7.3 Summary of the average in-sample and out-of-sample accuracy results from test 1.

Part	Training goal	In-sample accuracy, %		Out-of-sample accuracy, %	
		Mean	Standard Deviation	Mean	Standard Deviation
a	0.07	74.00	0.46	54.05	7.70
b	0.05	75.45	0.42	55.95	4.09
c	0.01	77.82	0.12	63.89	6.79
d	0.005	77.93	0	58.33	4.80
e	0.001	77.93	0	59.52	4.49

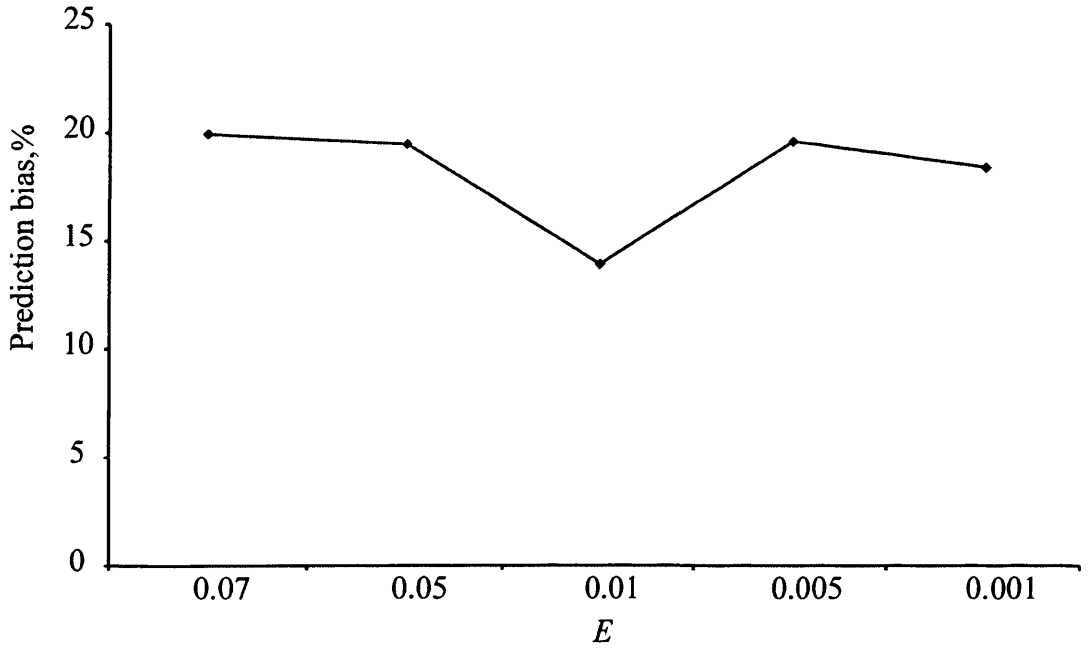


Figure 7.3 Prediction bias, % for different training goals, *E*.

As the training goal is decreased, that is the ANN is trained for a longer period, the average in-sample accuracy of the ANNs increases. However, this trend is not reflected in the out-of-sample results. Initially, as the training goal decreases the out-of-sample accuracy increases. At a training goal of 0.01 the out-of-sample accuracy reaches its maximum value. Beyond this point, although the in-sample accuracy increases, the out-of-sample accuracy decreases, that is, the ANNs begin to over-generalise. This is reflected in Figure 7.3 in which the prediction bias reaches its minimum value at a training goal of 0.01 and thereafter increases. The variability of the in-sample accuracy is seen to increase as the training goal is decreased, reflected in the decreasing value of the standard deviation. Conversely, the variability of the out-of-sample accuracy shows no particular trend. Comparing the standard deviation of the in-sample accuracy with the out-of-sample accuracy shows that the out-of-sample results are substantially more variable.

7.2.1.2 Test 2 results – Effect of changing the learning rate on classification accuracy

The results of test 2 to establish the effect of the learning rate on the in- and out-of-sample performance of the ANN classifiers are recorded in Appendix C (Tables C.6 to C.8 respectively). The average and standard deviation for the in- and out-of-sample accuracies are summarised in Table 7.4. The prediction bias for increased learning rate is depicted in Figure 7.4.

Table 7.4 Summary of the average in- and out-of-sample accuracy results from test 2.

Part	Learning Rate	In-sample accuracy, %		Out-of-sample accuracy, %	
		Mean	Standard Deviation	Mean	Standard Deviation
a	0.02	77.75	0.13	60.24	4.05
b	0.1	77.82	0.12	63.89	6.79
c	0.6	77.78	0.14	56.90	6.39

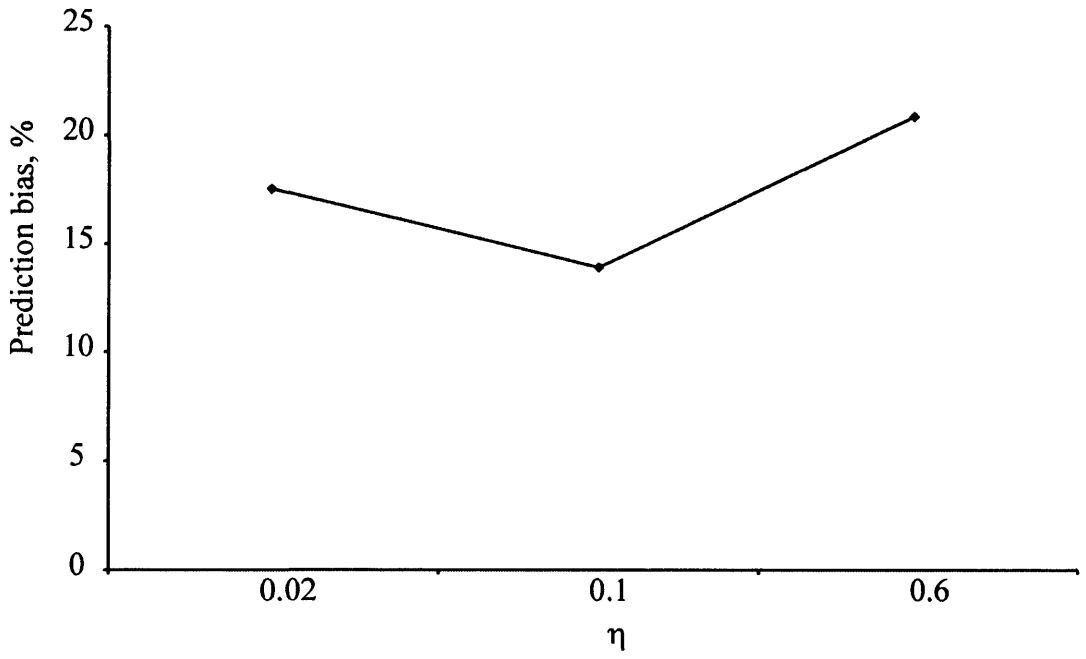


Figure 7.4 Prediction bias, % for different ANN learning rates, η.

Altering the value of the learning rate has a greater effect on the out-of-sample accuracy than on the in-sample accuracy. This can be seen in the larger range of average values reported for the out-of-sample than the in-sample accuracies. As the learning rate of the ANNs was increased the average in- and out-of-sample accuracy results showed similar trends. In increasing the learning rate from 0.02 to 0.1, both the average in- and out-of-sample accuracies increased. A further increase in the learning rate from 0.1 to 0.6 resulted in a decrease in both the average in- and out-of sample accuracies. As the learning rate is increased, the standard deviations of the in-sample and out-of-sample accuracies show opposing trends. The magnitudes of the standard deviations for the out-of-sample accuracies are an order greater than for the in-sample accuracies. The smallest prediction bias is obtained with a learning rate of 0.1.

7.2.1.3 Test 3 results – Effect of changing the momentum constant on classification accuracy

The results of test 3 to establish the effect of the momentum constant on performance are recorded in Appendix C (Tables C.9 and C.10 respectively). The average and standard deviation for the in- and out-of-sample accuracies are summarised in Table 7.5. The prediction bias is shown in Figure 7.5.

Table 7.5 Summary of the average in- and out-of-sample accuracy results from test 3.

Part	Momentum Constant	In-sample accuracy, %		Out-of-sample accuracy, %	
		Mean	Standard Deviation	Mean	Standard Deviation
a	0.9	77.82	0.12	63.89	6.79
b	0.8	77.80	0.11	56.90	7.73

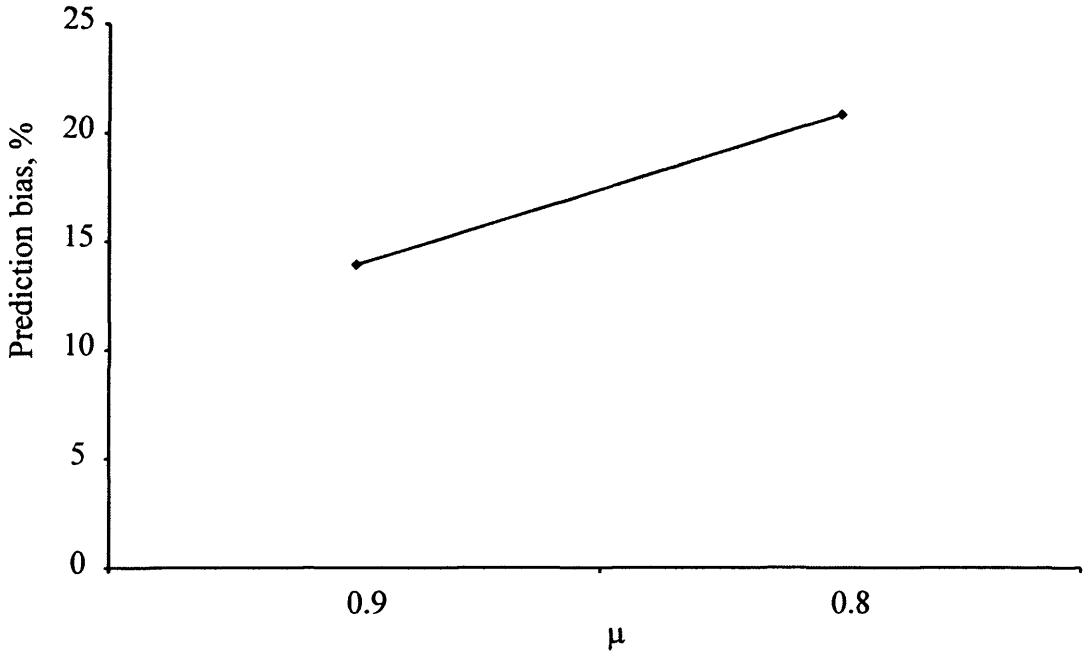


Figure 7.5 Prediction bias, % for different ANN momentum constants, μ.

As in the case of the learning rate, changing the value of the momentum constant has a greater effect on the out-of-sample accuracy than on the in-sample accuracy. This is reflected in the greater range of mean values reported for the out-of-sample than for the in-sample accuracies. As the momentum constant is decreased from 0.9 to 0.8 both the in-sample and out-of-sample accuracies decrease. The difference between the in-sample and out-of-sample accuracies increases. This difference is smallest for a momentum constant of 0.9. The standard deviation values are greater for the out-of-sample accuracies than the in-sample accuracies.

7.2.1.4 Test 4 results – Effect of changing the number of hidden neurons on classification accuracy

The results of test 4 to establish the effect of the number of hidden neurons on performance are recorded in Appendix C (Tables C.11 to C.14 respectively). The average and standard deviation for the in- and out-of-sample accuracies are summarised in Table 7.6. The prediction bias when the number of hidden neurons is increased is given in Figure 7.6.

Table 7.6 Summary of the average in- and out-of-sample accuracy results from test 4.

Part	Number of Hidden Neurons	In-sample accuracy, %		Out-of-sample accuracy, %	
		Mean	Standard Deviation	Mean	Standard Deviation
a	5	77.82	0.12	63.89	6.79
b	10	77.82	0.11	57.38	5.20
c	15	77.90	0.04	56.67	4.99
d	20	77.91	0.03	55.24	6.02

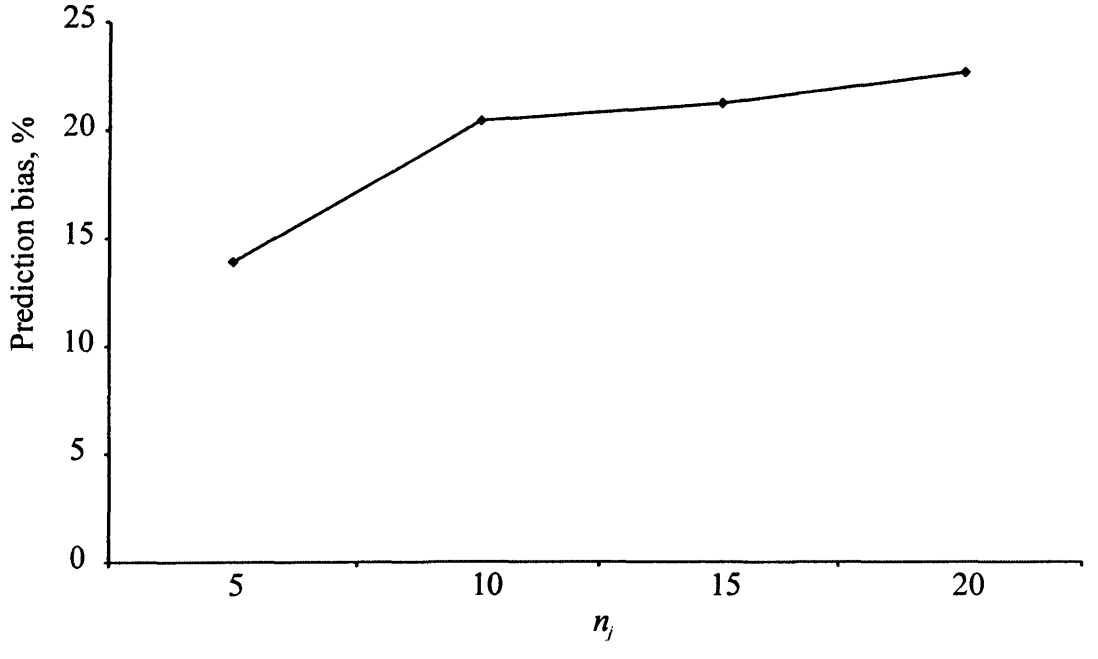


Figure 7.6 Prediction bias, % for different number of hidden neurons, n_j.

Increasing the number of hidden neurons had a greater effect on the mean out-of-sample accuracy than on the in-sample accuracy. Whilst the in-sample accuracy increased slightly with the increasing number of hidden neurons the out-of-sample accuracy decreased considerably. The prediction bias was at its minimum when 5 hidden neurons were utilised. As for the previous tests the standard deviation values are an order of magnitude greater for the out-of-sample accuracies than the in-sample accuracies.

7.2.1.5 Test 5 results – Effect of changing the number of output neurons on classification accuracy

The results of test 5 to establish the effect of changing the number of output neurons from 1 to 2 on performance are recorded in Appendix C (Tables C.15 and C.16 respectively). The average and standard deviation for the in- and out-of-sample accuracies are summarised in Table 7.7. The prediction bias is depicted in Figure 7.7.

Table 7.7 Summary of the average in-sample and out-of-sample accuracy results from test 5.

Part	Number of Output Neurons	In-sample accuracy, %		Out-of-sample accuracy, %	
		Mean	Standard Deviation	Mean	Standard Deviation
a	1	77.82	0.12	63.89	6.79
b	2	77.41	0.28	60.48	7.71

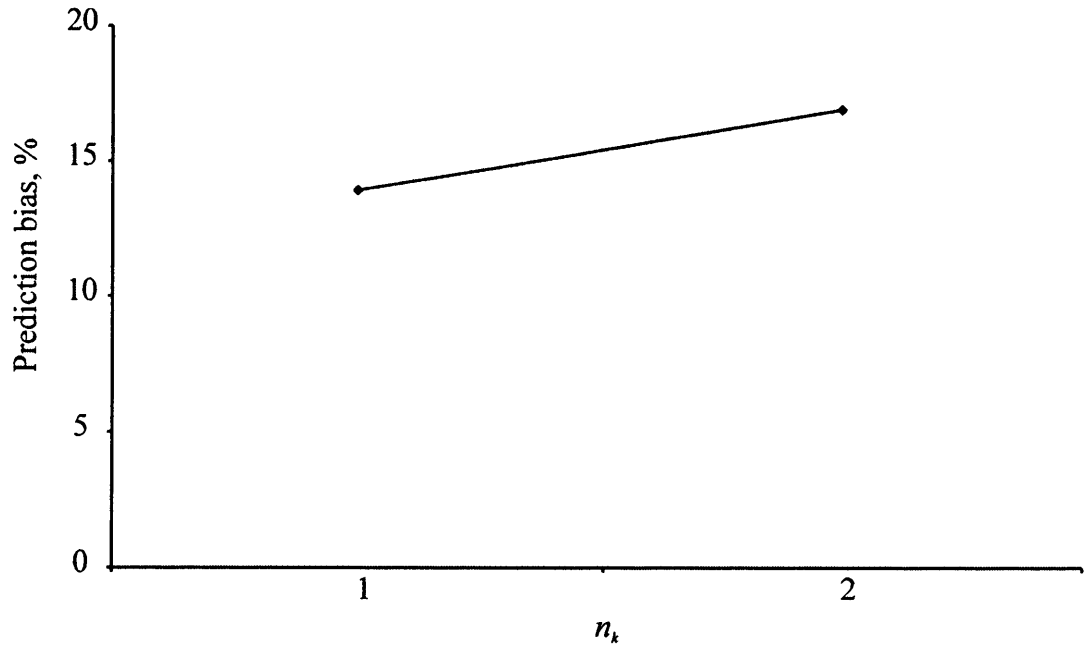


Figure 7.7 Prediction bias, % for different number of output neurons, n_k.

The addition of a second output neuron decreased both the average in- and out-of-sample accuracies, but, as noted in previous tests, had a greater effect on the out-of-sample accuracy than the in-sample accuracy. For both in- and out-of-sample cases, the variability in the accuracies is greater for the ANNs with two output neurons than the ANNs with a single output neuron. As before, the variability associated with the out-of-sample accuracy is greater than that with the in-sample accuracy. The prediction bias was smallest for the ANNs with one output neuron.

7.2.1.6 Summary of ANN accuracy results

The results of these five tests indicate that the best ANN configuration is an ANN with 5 hidden neurons and one output neuron trained using the following parameters: a learning rate of 0.1, a momentum rate of 0.9 and a MSE training goal of 0.01. This configuration produced the highest out-of-sample accuracy and the lowest prediction bias of all the ANNs. Using this configuration, the ANNs were able to classify subjects with an average in-sample accuracy of 77.82% and an average out-of-sample accuracy of 63.89%. The out-of-sample accuracy of the best ANN in this study is lower than the values reported in the literature, which lie in the range of 73–92% (Holzreiter and Köhle; Barton and Lees, 1995; Gioftsos and Grieve, 1995; Barton and

Lees, 1997; Lafuente *et al.*, 1997; Su and Wu, 2000; Wu and Su, 2000). Lafuente *et al.* (1997) and Wu and Su (2000) performed pre-selection of input variables using feature selection algorithms and thus tailored the use of the ANN more specifically to their applications. Lafuente *et al.* (1997) reported that the accuracy of the ANNs in their study varied greatly depending on the number of input variables used. The high higher accuracies reported by Barton and Lees (1995; 1997) were achieved using a small testing test; and those by Holzreiter and Köhle (1993) using a very large training set.

7.2.2 ANN Interpretability Results

The relative contribution of each variable (Table 7.8) to the final classification of subjects was examined for the optimum ANN configuration. The relative weights were calculated 42 times (for every left-out-person). The variables were ranked according to the absolute relative strength (equation (7.9)) where a variable with a ranking of 1 is deemed more important than a variable with a ranking of 2 etc. This process was repeated for each of the 10 runs. The rank of the variables based on the 420 iterations is recorded in Table 7.9. These results will be discussed in section 7.5.

Table 7.8 List of the variables, v_i (i = 1:18) used in the classification process

Variable, v_i	Variable Description
v_1	BMI
v_2	Cadence
v_3	Stance
v_4	APFPC1 Score
v_5	APFPC2 Score
v_6	APFPC3 Score
v_7	VFPC1 Score
v_8	VFPC2 Score
v_9	VFPC3 Score
v_{10}	FERPC1 Score
v_{11}	FERPC2 Score
v_{12}	AARPC1 Score
v_{13}	AARPC2 Score
v_{14}	AARPC3 Score
v_{15}	IERPC1 Score
v_{16}	ML Width
v_{17}	AP Width
v_{18}	Thigh Girth

Table 7.9 Ranking of input variables v_i (i = 1:18) from 10 training repetitions

Rank	Variable, v_i
1	v_8
2	v_2
3	v_4
4	v_{11}
5	v_{17}
6	v_{16}
7	v_{15}
8	v_{14}
9	v_3
10	v_6
11	v_5
12	v_1
13	v_7
14	v_{18}
15	v_9
16	v_{13}
17	v_{10}
18	v_{12}

7.3 LINEAR DISCRIMINANT ANALYSIS (LDA)

LDA was introduced in Chapter 5 in the context of feature selection. Here its application to classification is discussed.

Using LDA, classification functions are created for each group of the form

$$S_j = c_{j0} + c_{j1}v_1 + c_{j2}v_2 + \ldots + c_{jp}v_p \tag{7.10}$$

where j is the group, S_j the classification score for group j, v_i the input variable (i = 1: p), c_{j0} a constant term for group j and c_{ji} the classification function coefficient for group j. In matrix form, the classification function coefficients for group j are calculated using the formula

$$\mathbf{C}_j = \mathbf{VC}_w^{-1}\mathbf{M}_j \tag{7.11}$$

and the constant term as

$$c_{j0} = \left(-\tfrac{1}{2}\right)\mathbf{C}_j\mathbf{M}_j \tag{7.12}$$

where $\mathbf{VC}_w$ is the within-groups variance-covariance matrix, $\mathbf{M}_j$ the matrix of means of the p variables for group j and $\mathbf{C}_j$ the matrix of classification coefficients c_{ji} (Tabachnick and Fidell, 1989).

In this classification study of NL and OA knee function, subjects are assigned to a particular group (j = OA, NL) based on their classification scores, S_{OA} and S_{NL}. If $S_{OA} > S_{NL}$ then a subject is assigned to the OA group and vice-versa.

In addition to the classification functions LDA yields a discriminant function of the form

$$D = d_0 + d_1v_1 + d_2v_2 + \ldots + d_pv_p \tag{7.13}$$

where D is the discriminant score, d_0 a constant term and d_i the discriminant function coefficient. Once standardised, these coefficients can be examined to see the relative contribution that each variable makes to the discriminant function.

7.4 RESULTS OF THE LDA CLASSIFIER

A LDA procedure was implemented in SPSS 11.0 (SPSS Inc.). All variables (Table 7.8) were entered into the analysis simultaneously. Using all of the subjects, the LDA analysis yielded two classification functions and a discriminant function. The coefficients of these functions are recorded in Table 7.10 and 7.11 respectively.

7.4.1 LDA Accuracy Results

Using the classification coefficients recorded in Table 7.10, SPSS 10.0 automatically produced two classification scores for each subject and used these coefficients to assign the subjects to one of the two classes. The LDA procedure was able to classify subjects with an accuracy of 100%. Following this SPSS 10.0 used a LOOCV approach to classify each case using classification functions derived from all cases other than that case. The LDA procedure was able to classify these cross-validated cases with an accuracy of 95.20%. The out-of-sample accuracy is slightly greater than the values reported in the literature which are in the range 68–92% (Gioftsos and Grieve, 1995; Wu and Su, 2000).

7.4.2 LDA Interpretability Results

Standardising the coefficients of the discriminant function can show the relative importance of the input variables. This was determined using the standardised discriminant function derived from all cases as recorded in Table 7.12. These results will be discussed in section 7.5.

Table 7.10 Classification function coefficients c_{ji} for the input variables v_i (i = 1:18) and the constant terms c_{j0}, for the two groups (j = NL, OA)

Variable, v_i	**Coefficient, c_{ji}**	
	j = NL	**j = OA**
v_1	0.154	−0.971
v_2	15.453	14.007
v_3	112.045	108.927
v_4	10.356	10.513
v_5	−1.121	−0.563
v_6	−7.860	−6.950
v_7	12.938	12.817
v_8	1.066	2.315
v_9	21.385	20.278
v_{10}	−10.566	−10.416
v_{11}	−7.154	−6.249
v_{12}	3.238	3.149
v_{13}	0.229	0.468
v_{14}	9.406	8.368
v_{15}	0.324	0.731
v_{16}	−50.486	−38.736
v_{17}	−14.657	−11.212
v_{18}	9.471	8.676
Constant term, c_{j0}	−3684.812	−3525.264

Table 7.11 Discriminant function coefficients d_i for the input variables v_i (i = 1:18) and the constant term d_0

Variable, v_i	**Coefficient, d_j**
v_1	−0.162
v_2	−0.208
v_3	−0.449
v_4	0.023
v_5	0.080
v_6	0.131
v_7	−0.017
v_8	0.180
v_9	−0.160
v_{10}	0.022
v_{11}	0.130
v_{12}	−0.013
v_{13}	0.034
v_{14}	−0.149
v_{15}	0.059
v_{16}	1.693
v_{17}	0.496
v_{18}	−0.115
Constant term, d_0	23.156

Table 7.12 Standardised discriminant function coefficients ranked in descending order of absolute size of coefficient

Variable, v_i	**Coefficient, d_i**
v_{16}	2.159
v_1	−1.243
v_2	−0.985
v_3	−0.723
v_9	−0.582
v_{18}	−0.57
v_8	0.567
v_{17}	0.490
v_{14}	−0.452
v_{15}	0.440
v_{11}	0.406
v_5	0.395
v_6	0.357
v_{10}	0.165
v_{13}	0.149
v_7	−0.116
v_4	0.108
v_{12}	−0.098

7.5 DEMPSTER-SHAFER (DS) CLASSIFIER

The accuracy of the DS classifier is investigated in Chapter 4. In summary, the best DS classifier was able to classify subjects with an in-sample and out-of-sample accuracy of 97.62%.

Following Jones *et al.* (2004) a ranking of the variables was undertaken by means of an objective function (OB_{rank}). For a given input variable, the OB_{rank} is defined as the Euclidean distance of the mean coordinates of the two groups of subjects to their correct vertex in the simplex plot. As such it is a measure of the level of certainty that the input variable offers to the classification of subjects to their assigned class. Since in this thesis the sides of the simplex plot are of unit length ($0 \leq OB \leq 1$) (see section 3.4.4), a value of OB closer to zero implies a more robust classification. Conversely, a value of OB close to unity infers that the mean points are nearer to the Θ vertex or the other incorrect vertex. Consequently the input variable with the lowest associated OB was given a ranking of 1, the next lowest a ranking of 2 etc. For each variable the OB was calculated from the in-sample population. Using a LOOCV approach this was repeated 42 times. The final ranking of the input variables was then based on the average ranking from these 42 runs. The results of this ranking are given in Table 7.13.

The ranking of variables can also be visualised using the simplex plot as shown in Figure 7.8. The mean simplex coordinate of the in-sample subjects for each input variable lies near to the vertex associated with Θ. This is a direct consequence of the values assigned to the uncertainty limits ($[\Theta_L, \Theta_U] = [0.8,1]$). The horizontal line at the top of the simplex plot represents the limit at which $m(\Theta) = 0.8$. All of the individual variables have an associated level of uncertainty that is greater than 0.8. The section of the simplex plot highlighted in Figure 7.8a has been magnified in Figure 7.8b. As a consequence of using the above definition of OB_{rank}, the input variables that are given a high ranking by the DS classifier (e.g. v_2, v_8, v_4) are those whose simplex coordinates are situated at the furthest distance away from the uncertainty vertex. These variables, therefore, offer more weight to the final classification of subjects to their assigned group. In contrast, the simplex coordinates of the variables identified as least important (e.g., v_9, v_7, v_{10}) are positioned very near

to the uncertainty vertex and as such have less influence on the final classification of the subjects to their assigned group.

Table 7.13 Ranking of variables based on in-sample OB_{rank} of variables from 42 training repetitions, v_i (i =1:18)

Rank	Variable, v_i
1	v_8
2	v_2
3	v_4
4	v_{11}
5	v_{15}
6	v_{12}
7	v_{14}
8	v_3
9	v_1
10	v_6
11	v_{18}
12	v_{16}
13	v_{17}
14	v_5
15	v_{13}
16	v_{10}
17	v_7
18	v_9

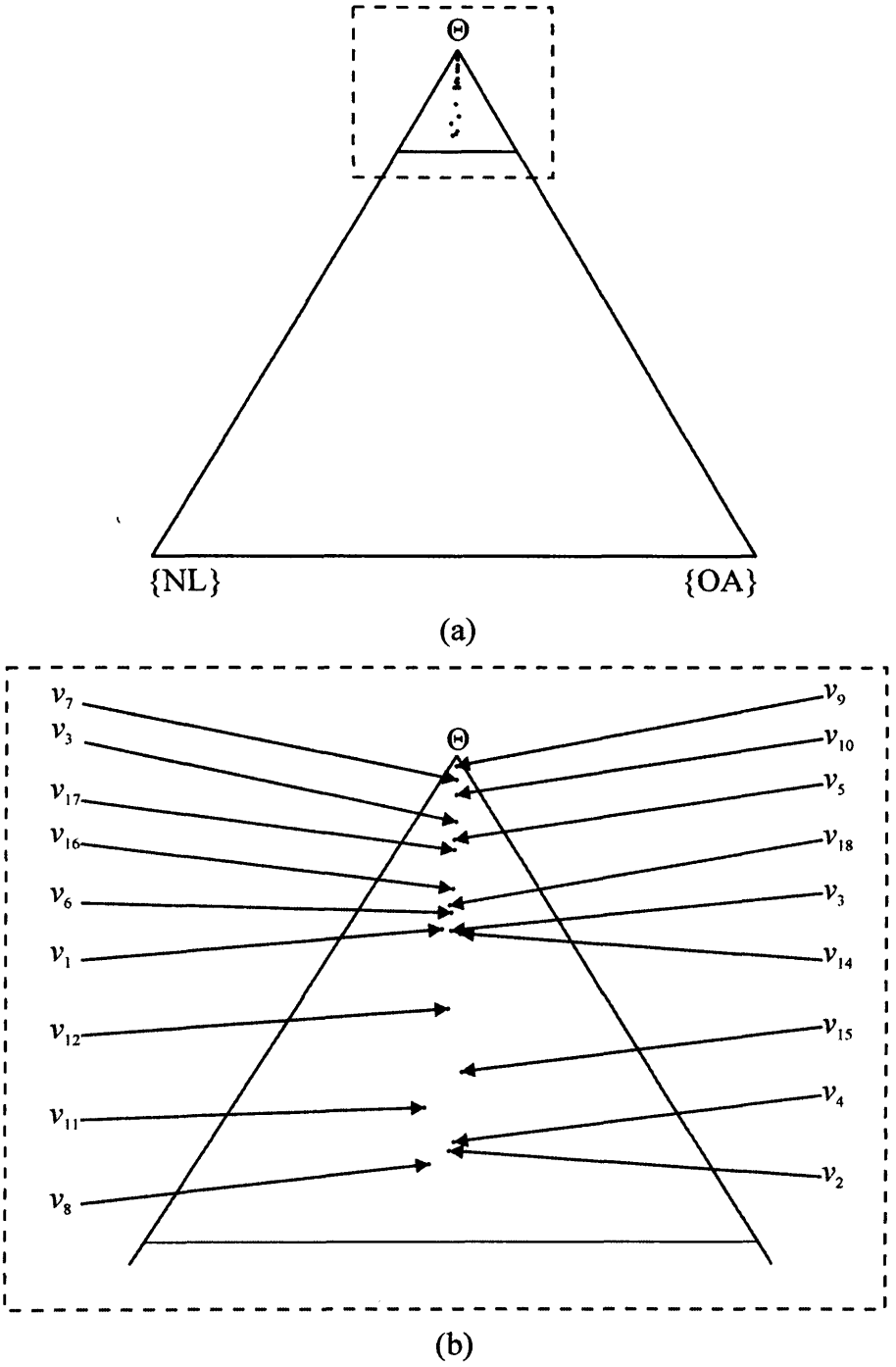


Figure 7.8 An example of the ranking of variables for one training repetition is given in terms of the positioning of the average simplex coordinate (from 41 subjects) for each variable, v_i (i = 1:18) in the (a) simplex plot and (b) an enlarged portion of the simplex plot highlighted by the dashed line in (a).

7.6 DISCUSSION OF RESULTS

This study was conducted in order to compare the performance of the DS classifier with two other classifiers namely an ANN classifier and an LDA classifier. Following Breiman (2001) the comparison of performance is based around two main areas: **prediction** and **interpretability**. In this study, prediction was measured in terms of the out-of-sample classification accuracy and interpretability was quantified using a ranking of variables. The results of this study have been presented in sections 7.2, 7.4 and 7.5 and shall now be discussed.

The in-sample and out-of-sample accuracy results for each classifier are summarised in Table 7.14. Comparing the DS classifier and the ANN classifier, the results show that the DS classifier has a higher in-sample and out-of-sample accuracy than the ANN classifier. The DS classifier has a lower in-sample classification accuracy than the LDA classifier but a higher out-of-sample accuracy. These results suggest that the DS classifier is superior to both the linear LDA classifier and the non-linear ANN classifier in terms of out-of-sample classification accuracy (prediction).

Table 7.14 Comparison of the in-sample and out-of-sample accuracy, %, of the DS, ANN and LDA classifiers

Classifier	In-sample accuracy, %	Out-of-sample accuracy, %
DS	97.62	97.62
ANN	77.82	63.89
LDA	100	95.24

The ten most important variables (in terms of ranking) identified by the three classifiers are given in Table 7.15. Here a ranking of 1 is given to the most important variable. The DS and ANN classifier identified similar variables as being important. Eight of the ten most important variables [v_2, v_3, v_4, v_6, v_8, v_{11}, v_{14}, v_{15}] were the same for these two classifiers. The DS and LDA identified the common variables [v_1, v_2, v_3, v_8, v_{14}, v_{15}] as important. All of these variables are often cited as being clinically

relevant. In terms of the nine variables identified as important by two or more of the classifiers a number of clinical observations will be discussed.

Table 7.15 Comparison of the ten most important variables, v_i identified by the DS, ANN and LDA classifiers

Classifier	Rank									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
DS	v_8	v_2	v_4	v_{11}	v_{15}	v_{12}	v_3	v_{14}	v_1	v_6
ANN	v_8	v_2	v_4	v_{11}	v_{17}	v_{16}	v_{15}	v_{14}	v_3	v_6
LDA	v_{16}	v_1	v_2	v_3	v_9	v_{18}	v_8	v_{17}	v_{14}	v_{15}

The variable v_8 was ranked first by both the DS classifier and the ANN classifier and ranked 8 by the LDA classifier. Table 7.8 shows that v_8 is the VFPC2 Score. From section 2.5.2.2 this is related to the vertical ground reaction force (GRF) (normalised to body weight) during the phase from loading response to mid-stance, namely the period from heel strike transient to the first peak vertical GRF (see Figure 2.18, Chapter 2, section 2.3). This period has been reported in the literature to be an important factor in discriminating OA and NL knee function. Gök *et al.* (2002) found that the first peak vertical GRF relative to body weight was significantly lower for subjects with OA knee function than for those with NL knee function. Childs *et al.* (2004) also reported that the maximum vertical GRF relative to body weight during loading was significantly lower in OA subjects than in NL subjects. Astephen *et al.* (2002a, 2002b) identified the loading response phase of the gait cycle to be an important factor in knee OA. Using the kinetic data collected during the clinical trial (see Chapter 2, section 2.3) an independent samples t-test (SPSS 11, SPPS Inc.) revealed that the peak vertical GRF was significantly lower (p = 0.05) in the OA population of subjects than in the NL population as shown in Table 7.16. Many OA subjects walk with an antalgic gait pattern since it is a natural response to reduce the loading through the painful knee.

Table 7.16 Independent t- test to identify significant differences between the OA and NL group means of the kinetic, kinematic, temporal-distance and anthropometrical parameters

Variable	Mean		Significance
	OA Sample	NL Sample	
Peak vertical GRF	1.0688	1.2076	0.006
Cadence (min^{-1})	43.5688	53.9382	0.000
Peak anterior GRF	0.1141	0.2321	0.000
Peak posterior GRF	-0.0918	-0.1936	0.000
Peak flexion (°)	48.2163	56.4752	0.000
Peak internal rotation (°)	-5.1296	-10.780	0.002
Stance Phase (% gait cycle)	62.0007	59.8966	0.000
Peak adduction (°)	-4.8123	-2.8383	0.134
Peak abduction (°)	6.4785	7.1739	0.673
BMI (kgm^{-2})	28.5509	26.1222	0.311

The variable v_2 was ranked second by both the DS classifier and the ANN classifier and third by the LDA classifier. From Table 7.8 v_2 is the variable cadence, which has been reported in the literature to be a parameter that differs for the OA population compared to the NL population. Gök *et al.* (2002) reported the cadence of OA subjects to be significantly less than in NL subjects. From the temporal-distance data collated as part of the clinical knee trial (see Chapter 2, section 2.4) an independent samples t-test (SPSS 11, SPPS Inc.) found a significant difference in the mean values of the cadence for the OA and NL group ($p < 0.001$). The lower cadence is attributed to an increase in the time taken for one gait cycle which results in a lengthening of the time spent in double support rather than single support (Whittle, 1996). This is a compensatory mechanism to reduce pain as it reduces the time that the subject spends on the affected leg in single support. However, it must be recognised that a difference in cadence between the two groups of subjects may not be uniquely attributable to a

presence of knee OA. The presence of co-morbidities such as angina and shortness of breath may also be contributing factors, especially in the more elderly subjects.

The variable v_4 was ranked third by the DS and ANN classifiers. This variable is the APFPC1 Score (Table 7.8). From section 2.5.2.1 this variable is related to the peak anterior force and the peak posterior GRF (normalised to body weight). The tenth ranked variable by the DS and ANN classifiers is v_6, the APFPC3 Score. From section 2.5.2.1 this variable is related to the anterior-posterior GRF (normalised to body weight) during late pre-swing. Schneider and Chao (1983) reported that the magnitude of the anterior-posterior GRF was less in subjects with knee joint disease than in normal subjects. Using the kinetic data collated as part of the clinical knee trial (see Chapter 2, section 2.3) an independent samples t-test (SPSS 11, SPPS Inc.) found that both the peak anterior GRF and the peak posterior GRF were significantly lower ($p < 0.001$) in the OA population of subjects than in the NL population. The behaviour of the anterior-posterior GRF during level walking is related to the movement of the centre of mass (Meglan and Todd, 1994). During loading response the body's centre of mass in the sagittal plane lies at a point behind the centre of pressure and is accelerating in an anterior direction. This requires that the anterior-posterior GRF act in a posterior direction to oppose the force. As the body passes over the supporting foot in late stance, the centre of mass moves ahead of the centre of pressure and accelerates in a posterior direction. Consequently, the anterior-posterior GRF acts in an anterior direction to oppose this force. As a subject's walking speed is reduced, as evident in subjects with OA knee function, the acceleration of the centre of mass in both anterior and posterior directions decrease. Consequently the peaks of the anterior-posterior GRF are reduced. In addition to this is seems obvious, as with the vertical GRF, that as a protective mechanism and with associated compromised muscle and soft-tissue function, the subject with knee OA would place less force through the painful knee.

Table 7.8 shows that the variable ranked fourth by the DS classifier and the ANN classifier, v_{11} is the FERPC2 Score. From section 2.5.2.3 the FERPC2 Score is related to the knee flexion during the phase from 58% to 76% of the gait cycle. Figure 2.13 (Chapter 2, section 2.2.3) shows that during this period of the swing phase the normal knee obtains peak flexion. Kaufman *et al.* (2001) reported a significant difference of 6

degrees in the peak knee flexion of subjects with OA function and subjects with NL function. Al-Zahrani and Bakheit (2002) found that the peak knee flexion during swing phase was significantly less for OA subjects than for NL subjects. Gök *et al.* (2002) found a significant difference of 9 degrees in the maximum knee flexion during swing between OA and NL subjects and attributed this reduction in the range of motion to a decrease in joint flexibility. Using the kinematic data collected during the clinical trial (see Chapter 2, section 2.2) an independent samples t-test (SPSS 11, SPPS Inc.) revealed that the peak flexion was significantly lower ($p < 0.001$) in the OA population of subjects than in the NL population. In the OA knee the level of deformity dictates the movement of the joint. The combination of muscle atrophy, soft tissue contracture and an increased congruency of the joint surfaces results in a reduction in the range of motion of the knee joint. Additionally, it has been reported that the range of knee flexion is highly dependent on walking speed. During swing phase, the leg acts as a double-jointed swinging pendulum (Whittle, 1996). It follows then that at slower speeds the momentum of the leg would decrease resulting in a reduction in the knee range of motion.

The fifth ranked variable by the DS classifier is v_{15}, the IERPC1 Score as shown in Table 7.8. This variable was ranked seventh and tenth by the ANN and LDA classifiers respectively. From section 2.5.2.5 the IERPC1 Score is related to the internal-external rotation waveform from loading response to mid swing. Holt *et al.* (2002) reported a difference between the internal-external rotational range of motion of subjects with NL and OA knee function. Using the kinematic data collected during the knee clinical trial (see Chapter 2, section 2.2) an independent samples t-test (SPSS 11, SPPS Inc.) showed that that the peak internal rotation was significantly lower ($p < 0.05$) in the OA population of subjects than in the NL population. As for the flexion-extension movement the reduction in internal rotation may be attributed to a stiffening of the joint. In addition the internal-external range of motion is affected by the flexion-extension movement of the knee joint.

The seventh ranked variable by the DS classifier is v_3, the percentage stance phase. This variable was also identified as important by the ANN and LDA classifiers. Al-Zahrani and Bakheit (2002) found that the stance phase of the gait cycle was more prolonged in subjects with OA knee function compared to those with NL function.

Similarly Gök *et al.* (2002) reported as overall stance time that was longer in the OA group. From the temporal-distance data collected during the clinical trial (see Chapter 2, section 2.4) an independent samples t-test (SPSS 11, SPPS Inc.) showed that the stance phase (as a percentage of the gait cycle) was significantly longer ($p < 0.001$) in the OA population of subjects than in the NL population. Lengthening the stance phase increases the period of double support (Whittle, 1996). This is another compensatory mechanism to help pain reduction since it limits the time that the subject spends on the affected limb in single support. This length of stance phase is also affected by the speed of walking

The variable ranked eighth by the DS and ANN classifiers and ninth by the LDA classifier is v_{14}, the AARPC3 Score. From section 2.5.2.4 this variable is related to the Abduction-Adduction rotation during terminal swing. Gök *et al.* (2002) found a significant difference in the maximum knee abduction angle in swing between the group of OA subjects and the group of NL subjects. However, using the kinematic data collected during the clinical trial (see Chapter 2, section 2.2) an independent samples t-test (SPSS 11, SPPS Inc.) found that although a difference was observed in the mean abduction angle between the OA and NL subjects, this difference was not found to be significant ($p > 0.05$). Additionally the peak adduction angle was less in the NL population than in the OA. However, this difference was not significant ($p > 0.05$). The difference in the abduction-adduction angle may be attributed to deformity of the joint.

The variable ranked in ninth position by the DS classifier and second by the LDA classifier, v_1 is the BMI. Obesity or an increased BMI are seen as a risk factor for knee OA (Foye *et al.*, 2000; Nevitt and Lane, 1999; Creamer and Hochberg, 1997; Gelber *et al.*, 1999). Additionally, Kaufman *et al.* (2001) reported that the gait patterns of subjects with OA knee function that also had an increased BMI, deviated further from normal than those with a low BMI. Using the anthropometric data collected during the knee clinical trial (see Chapter 2, section 2.4) an independent samples t-test (SPSS 11, SPPS Inc.) revealed that although the mean BMI was greater in the OA population of subjects compared to the NL population, this difference was not found to be significant ($p > 0.05$). A question exists in the literature as to whether an increased BMI is a precursor to knee OA or whether knee OA leads to an increased BMI

(Gelber *et al.*, 1999). Both Gelber *et al.* (1999) and Nevitt and Lane (1999) comment that an increase in BMI leads to an increase in knee joint loading which can result in knee OA. Conversely, Gelber *et al.* (1999) also suggest that subjects with knee OA may adopt a more sedentary lifestyle due to pain and limited mobility and as a result gain weight.

Although it is evident that the variables discussed above are independently clinical indicators of OA knee function the variables must be looked at collectively, as would be the practice in a clinical setting, in order to obtain a definitive OA diagnosis.

This chapter has compared the performance of the DS classifier with an ANN classifier and an LDA classifier. The results have been presented and a discussion based on these results given. Conclusions from this study and suggested further work are given in Chapter 8.

CHAPTER 8

CONCLUSIONS AND FURTHER WORK

8.1. CONCLUSIONS

Developments in gait analysis technology over the last two decades have enhanced our understanding of human locomotion. However, such advances in knowledge are futile if no practical use is made of them. Scientists and engineers need to make the most of these developments by forging stronger links with orthopaedic surgeons and applying further advances in their knowledge to clinical problems for the long-term benefit of patients.

Over the last few decades, this need has been identified by many in the field of biomechanics and biomedical engineering and a "serious attempt [has been made] to take gait analysis out of the research laboratory and into the clinic" (Whittle, 1996 pp.58). With this need as the driving-force, the aim of this research has been to develop an objective classification tool using motion analysis for a proposed application to clinical diagnostics and monitoring. The purpose of the tool is to aid orthopaedic surgeons and therapists in making clinical decisions and so to promote increased confidence in a patient's medical care.

An extensive database of normal (NL), osteoarthritic (OA) and total knee replacement (TKR) knee function data – i.e. kinematic, kinetic, temporal-distance and anthropometrical parameters – has been established during a clinical trial in Cardiff University. With a view to developing a diagnostic tool for OA progression and TKR recovery, initial studies found that the extensive range of data acquired was extremely difficult to analyse objectively and that an automated approach had to be found.

Emerging objective classification techniques, e.g. Artificial Neural Networks were found to be potentially useful tools for such a task. However, these techniques are of the "black box" approach and as such are often perceived as unappealing to the clinical community since they are difficult to realize mathematically and provide little

indication of the important factors used for classification. Furthermore, the most successful applications have necessitated a large set of training patterns, which is often not practically attainable in the clinical setting and not available in this study.

Faced with this problem, a diagnostic tool has been developed that is based around the Dempster-Shafer (DS) theory, and as such is built on the sound foundations of Bayesian statistics. Using the data collected during the clinical knee trial, this novel approach enables automated, objective classification of subjects into an OA or NL group. The kinematic and kinetic waveforms are pre-processed using Principal Component Analysis. Each piece of data is then transformed into a set of belief values: a level of belief that a subject has OA knee function, a level of belief that a subject has NL knee function and an associated level of uncertainty. These three belief values are subsequently represented on a simplex plot, which enables the final classification of a subject and the variables contributing to that classification to be represented visually, and the level of benefit achieved by TKR surgery to be quantified.

If this new tool is to provide an enhancement to diagnosis, orthopaedic intervention and rehabilitation, it will require confident use by orthopaedic surgeons, therapists and biomechanical engineers collectively. Consequently, the tool must be accurate and clinically relevant.

The results reveal that the DS method is a highly accurate classification tool. It was able to classify new subjects with an accuracy of 97.62%, compared to two well-established methods – the Artificial Neural Network and Linear Discriminant Analysis – which were able to classify subjects with accuracies of 63.89% and 95.24% respectively.

The clinical relevance and appeal of the tool is enhanced in two main ways: visualisation and the transparency of the contribution that the variables make to the overall classification.

The simple visual output shows the extent of a patient's pathology and recovery and gives an indication of those variables that are most influential in producing the classification. Results showed that the variables identified by the tool as the most

influential parameters in distinguishing OA and NL knee function are those that are often cited to be clinically relevant. This is a significant finding and is one aspect that has enhanced its appeal to the clinical community.

Finally, the DS classifier has been developed in such a way that it can be used as a fully or partially automated tool. For those with a clinical background the input variables can be chosen manually. With those who are not familiar with the clinical problem or in cases where it is not obvious as to which variables are most useful in determining a subject's class, the input variables can be selected using a feature selection algorithm. Additionally the DS control parameters, which are an intrinsic part of the tool, can be chosen by an expert or by an optimisation approach.

The initial results of this application have demonstrated a logical, practical and visual approach that can be used to differentiate between the characteristics of NL and OA knee function and to diagnose the extent to which a patient has developed OA and recovered after subsequent TKR surgery. Initial studies using this technique have provided encouraging results in terms of accuracy, validity and clinical relevance.

8.2 FURTHER WORK

In order to meet the overall aims of this project the novel diagnostic tool needs further development in a number of ways.

The mathematical foundation of the classification system is based on the DS theory and consequently allows a level of ignorance to exist throughout the analysis. The specific aspects of ignorance intrinsic to the new classification method in the context of motion analysis were defined in Chapter 3 as uncertainty, imprecision and incompleteness. In terms of imprecision, errors due to skin movement artefacts can result in inaccuracies in the rotational measurements of the knee joint in the order of 10 to 100% (Cappozzo *et al.*, 1996). Therefore, the specific errors associated with skin movement artifacts and the reconstruction of marker coordinates must be quantified and subsequently integrated into the classifier. In terms of incompleteness, methods for dealing with missing data, such as the medial-lateral ground reaction force data, must be investigated, e.g., the combinatorial model (see Smets, 1991).

Alternatively, in the case of missing variable, a body of evidence with $m(\Theta) = 1$ and $m(\{OA\}) = m(\{NL\}) = 0$ could be assigned.

The stability and accuracy, and hence the validity, of the classification method depend on a set of control parameters, k, θ, A and B, whose values are assumed by an expert or by use of an optimisation approach. Using the simulated annealing optimisation technique to assign values to the DS control parameters resulted in a decrease in the in-sample and out-of-sample accuracy. This was a direct consequence of the limits placed on the DS control parameters A and B. It would be desirable to use the optimisation technique because it removes the need for expert opinion, therefore, further investigation of the use of optimisation methods in the assignment of values to these DS control parameters is needed in particular the effect of the DS control parameters A and B on the accuracy of the classification method. This could be undertaken by repeating the optimisation tests using different limits on A and B, and by assigning different limits to each input variable (see Beynon, 2004). With the aim of developing a fully automated tool with high classification accuracy and a low associated level of uncertainty, it would be worthwhile to investigate the effect of changing the control parameters A and B on the performance of the feature selection algorithm by repeating the feature selection tests with different limits.

Differences were identified between the DS classifier outcomes, the subjective clinical opinions and the quality-of-life scores. The main reason for this inconsistency is that each of the three methods provides a different perspective on the assessment of knee function. The DST method assesses knee function during level walking. The quality-of-life scores measure knee function during different daily activities and consider clinical parameters such as buckling, instability and pain. The orthopaedic surgeon considered the static alignment of the knee, passive range of motion and the knee function whilst rising from a chair in addition to the knee function during level walking. The surgeon also considered what was achievable by the individual patient and measured improvement on these terms rather than by comparison with the NL population. Further work is required to study this aspect if a fair comparison between the DS classifier and the clinical and quality-of-life scores is to be obtained and to ensure that the objective DS classifier and subjective clinical opinions are consistent. This may be achieved by tailoring the input variables to suit the clinical question.

At various stages of the research, time was spent talking to orthopaedic surgeons in order to understand more about the clinical diagnosis of knee OA and to identify whether the objective tool makes decisions in line with subjective clinical opinion. This partnership has provided an invaluable insight in understanding the clinical application of the DS method. It is suggested that more time be spent in conjunction with orthopaedic surgeons and patients in clinics to investigate the feasibility of the use of the proposed diagnostic tool in a clinical setting.

The DS classifier is a generic method and, as such, it is applicable to a wide range of classification and predictive problems. The application of the DS classifier to other biomechanical and clinical problems should be investigated, e.g., identification of ankle, spine and hip pathology. Additionally, it is suggested that the method be used for the assessment of the relative merits of different treatment options, devices and surgery, e.g., identification of the differences between rotating platform and fixed bearing knee implants.

REFERENCES

Al-Zahrani, K.S. and Bakheit, A.M.O. (2002). A study of the gait characteristics of patients with chronic osteoarthritis of the knee. *Disability and Rehabilitation.* 24, 5, pp.275-280.

Andriacchi, T.P. (1993). Functional analysis of pre and post-knee surgery: total knee arthroplasty and ACL reconstruction. *Journal of Biomechanical Engineering.* 115, pp.575-581.

Andriacchi, T.P. and Alexander, E.J. (2000). Studies of human locomotion. *Journal of Biomechanics.* 33, pp.1217-1224.

Arthritis Research Campaign (1998). Osteoarthritis of the knee. [WWW]. <URL: http://www.arc.org.uk/about_arth/booklets/6027/OAofKnee.pdf [11 June 2004].

Astephen, J.L., Deluzio, K.J. and Wyss, U.P. (2002a). A simultaneous analysis of multiple waveform and constant gait measures: application to knee osteoarthritis. *Proceedings of the 4th World Congress of Biomechanics, Calgary, Canada.*

Astephen, J.L., Deluzio, K.J. and Wyss, U.P. (2002b). The loading response phase of the gait cycle is important to knee osteoarthritis. *Proceedings of the 4th World Congress of Biomechanics, Calgary, Canada.*

Baker, J.E. (1987). Reducing bias and inefficiency in the selection algorithm. 'In: *Genetic Algorithms and their Applications: Proceedings of the International Conference on Genetic Algorithms, Cambridge, MA*'. Grefenstette, J.J. ed. London: Lawerence Erlbaum Associates. pp.14-21.

Barton, J.G. and Lees, A. (1995). Development of a connectionist expert system to identify foot problems based on under-foot pressure patterns. *Clinical Biomechanics.* 10, 7, pp.385-391.

Barton, J.G. and Lees, A. (1997). An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams. *Gait and Posture*. 5, pp.28-33.

Benedetti, M.G., Catani, F., Leardini, A., Pignotti, E. and Giannini, S. (1998). Data management in gait analysis for clinical applications. *Clinical Biomechanics*. 13, 3, pp.204-215.

Benedetti, M.G., Catani, F., Bilotta, T.W., Marcacci, M., Mariani, E. and Giannini, S. (2003). Muscle activation pattern and gait biomechanics after total knee replacement. *Clinical Biomechanics*. 18, pp.871-876.

Beynon, M.J., Curry, B. and Morgan, P. (2000). The Dempster-Shafer theory of evidence: an alternative approach to multicriteria decision modelling. *Omega*. 28, pp.37-50.

Beynon, M.J. (2004). A novel technique of object ranking and classification under ignorance: an application to the corporate failure risk problem. *European Journal of Operational Research*. In Press.

Beynon, M.J., Jones, L., Holt, C.A. (2002). Classification of osteoarthritic and normal knee function using three dimensional motion analysis and the Dempster-Shafer theory of evidence. *Proceedings of the International Society of Biomechanics – 7th Symposium on Human Motion Analysis, Newcastle, UK, July*. pp.85-88. Podium presentation.

Bishop, C.M. (1995). *Neural networks for pattern recognition*. Oxford: Clarendon Press New York: Oxford University Press.

Borzelli, G., Cappozzo, A. and Papa, E. (1999). Inter- and intra-individual variability of ground reaction forces during sit-to-stand with principal component analysis, *Medical Engineering & Physics*. 21, pp.235-240.

Breiman, L. (2001). Statistical modelling: the two cultures. *Statistical Science*. 16, 3, pp.199-231.

Brigham, E. O. (1988). *The fast Fourier transform and its applications*. Englewood Cliffs, N.J.: Prentice Hall.

Brill, F.Z., Brown, D.E. and Worthy, N.M. (1992). Fast genetic selection of features for neural network classifiers. *IEEE Transactions on Neural Networks*. 3, 2, pp.324-328.

Callister, W.D. (1999). *Materials Science and Engineering: An Introduction*. 5th edition. New York; Chichester: Wiley.

Cappello, A., Cappozzo, A., Della Croce, U. and Leardini, A. (1997). Bone position and orientation reconstruction using external markers. In: *Three-dimensional analysis of human locomotion*. Eds. Allard, P., Cappozzo, A., Lundberg, A., Vaughan, C. Chapter 8, pp.146-171. John Wiley and Sons Ltd.

Cappozzo, A., Catani, F., Leardini, A., Benedetti, M.G. and Della Croce, U. (1996). Position and orientation in space of bones during movement: experimental artefacts. *Clinical Biomechanics*. 11, 2, pp.90-100.

Catani, F., Benedetti, M.G., De Felice, R., Buzzi, R., Giannini, S. and Aglietti, P. (2003). Mobile and fixed bearing total knee prosthesis functional comparison during stair climbing. *Clinical Biomechanics*. 18, pp.410-418.

Chao, E.Y., Laughman, R.K. and Stauffer, R.N. (1980). Biomechanical gait evaluation of pre and postoperative total knee replacement patients. *Archives of Orthopaedic and Traumatic Surgery*. 97, pp.309-317.

Chao, E.Y., Laughman, R.K., Schneider, E. and Stauffer, R.N. (1983). Normative data of knee joint motion and ground reaction forces in adult level walking, *Journal of Biomechanics*, 16, 3, pp.219-233.

Chau, T. (2001a). A review of analytical techniques for gait data. Part 1: fuzzy, statistical and fractal methods. *Gait and Posture*. 13, pp.49-66.

Chau, T. (2001b). A review of analytical techniques for gait data. Part 2: neural network and wavelet methods. *Gait and Posture*. 13, pp.102-120.

Childs, J.D., Sparto, P.J., Fitzgerald, G.K., Bizzini, M. and Irrgang, J.J. (2004). Alterations in lower extremity movement and muscle activation patterns in individuals with knee osteoarthritis. *Clinical Biomechanics*. 19, pp.44-49.

Chipperfield, A., Fleming, P., Pohlheim, H. and Fonseca, C. (1994). *Genetic algorithm toolbox user's guide, for use with Matlab*, v1.2. [WWW]. <URL: http://www.shef.ac.uk/~gaipp/ga-toolbox/manual.pdf [21 January 2004].

Chipperfield, A.J. and Fleming, P.J. (1995). The MATLAB genetic algorithm toolbox 'In: IEE Colloqium on Applied Control Techniques Using MATLAB, Digest No. 1995/014' [WWW]. <URL: http://www.shef.ac.uk/~gaipp/ga-toolbox/intro_paper_1.pdf [21 January 2004].

Comrey, A. L. (1973). *A first course in factor analysis*. New York; London: Academic Press

Coxeter, H.S.M. (1969). *Introduction to Geometry*, 2nd edition. London: Wiley. Chapter 13.7.

Creamer, P. and Hochberg, M.C. (1997). Osteoarthritis. *The Lancet*. 350, 16, pp.503-509.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*. 2, pp.303-314.

Dash, M. and Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis*. 1, pp.131-156.

Daultrey, S. (1976). Principal components analysis. Norwich: Geo Abstracts Ltd.

Davies, A.P. (2002). Rating systems for total knee replacement. *The Knee*. 9, pp.261-266.

Davis, B.L. and Vaughan, C.L. (1993). Phasic behaviour of EMG signals during gait: use of multivariate statistics. *Journal of Electromyography and kinesiology*. 3, 1, pp.51-60.

Deluzio, K.J., Wyss, U.P., Zee, B., Costigan, P.A. and Sorbie C. (1997). Principal component models of knee kinematics and kinetics: normal vs. pathological gait patterns. *Human Movement Science*. 16, pp.201-217.

Deluzio, K.J., Wyss, U.P., Costigan, P.A., Sorbie, C. and Zee, B. (1999). Gait assessment in unicompartmental knee arthroplasty patients: principal component modelling of gait waveforms and clincial status. *Human Movement Science*. 18, pp.701-711.

Dempster, A.P. (1968). A generalization of Bayesian inference (with discussion), *J Roy Stat Soc, Series* B. 30, 2, pp.205–247.

Denoeux, T. (1997). Analysis of evidence-theoretic decision rules for pattern classification. *Pattern Recognition*. 30, 7, pp.1095-1107.

Dowsland, K.A. (1995). Simulated annealing 'In: *Modern heuristic techniques for combinatorial problems*', Reeves, C.R. ed. London: McGraw-Hill.

Duda, R.O., Hart, P.E. and Stork, D.G. (2001). *Pattern Classification*. 2nd edition. New York; Chichester: Wiley.

Foye, P.M., Stitik, T.P., Chen, B. and Nadler, S.F. (2000). Osteoarthritis and body weight. *Nutrition Research*. 20, 6, pp.899-903.

Fuchs, S., Flören, M., Skwara, A. and Tibesku, C.O. (2002). Quantitative gait analysis in unconstrained total knee arthroplasty patients. *International Journal of Rehabilitation Research.* 25, pp.65-70.

Garson, G.D. (1991). Interpreting neural-network connection weights. AI Expert. April, pp.47-51.

Gelber, A.C., Hochberg, M.C., Mead, L.A., Wang, N-Y., Wigley., F.M. and Klag, M.J. (1999). Body mass index in young men and the risk of subsequent knee and hip osteoarthritis. *The American Journal of Medicine.* 107, pp.542-548.

Gerig, G., Welti, D., Guttmann, C.R.G., Colchester, A.C.F. and Szekely, G. (2000). Exploring the discrimination power of the time domain for segmentation and characterization of active lesions in serial MR data. *Medical Image Analysis.* 4, pp.31-42.

Gioftsos, G. and Grieve, D.W. (1995). The use of neural networks to recognize patterns of human movement: gait patterns. *Clinical Biomechanics.* 10, 4, pp.179-183.

Gök, H., Ergin, S. and Yavuzer, G. (2002). Kinetic and kinematic characteristics of gait in patients with medial knee arthrosis. *Acta Orthop Scand.* 73, 6, pp.647-652.

Green, P.E. (1978). *Analyzing multivariate data.* Hinsdale, Illinois: Dryden Press.

Grood, E.S. and Suntay, W.J. (1983). A joint co-ordinate system for the clinical description of three-dimensional motions: Application to the knee, *Journal of Biomechanical Engineering.* 105, pp.136-144.

Holt, C.A., Hayes, N.J., van Deursen R.W.M. and O'Callaghan, P.T. (2000). Three-dimensional analysis of the tibiofemoral joint using external marker clusters and the JCS approach – Comparison of normal and osteoarthritic knee function. *Computer Methods in Biomechanics and Biomedical Engineering 3.* Lisbon. Gordon and Breach Science Publishers SA. pp.289-294.

Holt, C.A., Jones, L., O'Callaghan, P.O., Roy, S. and Wilson, C. (2002). Comparison of fixed and rotating bearing knee replacements in-vivo using a three-dimensional motion capture technique: preliminary results. *Proceedings of the 4th World Congress of Biomechanics, Calgary, Canada.*

Holzreiter, S.H. and Köhle, M.E. (1993). Assessment of gait patterns using neural networks. *Journal of Biomechanics.* 26, 6, pp.645-65.

Irrgang J.J., Synder-Mackler, L., Wainner, R.S., Fu, F.H. and Harner, C.D. (1998). Development of a patient-reported measure of function of the knee. *The Journal of Bone and Joint Surgery.* 80-A, 8, pp.1132-1145.

Jacobs, N.A., Skorecki, J. and Charnley, J. (1972). Analysis of the vertical component of force in normal and pathological gait. *Journal of Biomechanics.* 5, pp.11-34.

Jain, A. and Zongker, D. (1997). Feature selection: evaluation, application and small sample performance. IEEE Transactions on Pattern Analysis and Machine Intelligence. 19, 2, pp.153-158.

Ji, Q. and Marefat, M.M. (2003). A Dempster-Shafer approach for recognizing machine features from CAD models. *Pattern Recognition.* 36, pp.1355-1368.

Jolliffe, I.T. (1986). Principal component analysis. New York: Springer.

Jones, L., Beynon, M.J. and Holt, C.A. (2003a). *Performance of a new gait classification method.* Proceedings of Salford's 2nd International Conference on "*Biomechanics of the Lower Limb in Health, Disease and Rehabilitation*", Salford, UK, October. Podium presentation. Best Abstract Award.

Jones, L., Beynon, M.J. and Holt, C.A. (2003b). How important is the choice of input variables when designing a novel classifier for normal and osteoarthritic knee function? *Proceedings of the International Congress on Computational Bioengineering,* Zaragoza, Spain, October. Podium presentation.

Jones L., Beynon M.J. and Holt C.A. (2003c) *An application of the Dempster-Shafer theory of evidence to the classification of knee function and detection of improvement due to total knee replacement surgery.* Accepted for publication in Journal of Biomechanics.

Jones L., Beynon M.J., Holt C.A., (2004) Two different gait classifiers – a comparative study, *Computer Methods in Biomechanics and Biomedical Engineering 6*, Madrid.

Kaufman, L. and Rousseeuw, P.J. (1990). Finding groups in data: an introduction to cluster analysis. New York: Wiley.

Kaufman, K.R., Hughes, C., Morrey, B.F., Morrey, M. and An, K-N. (2001) Gait characteristics of patients with knee osteoarthritis. *Journal of Biomechanics.* 34, pp.907-915.

Kendall, M.G. and Stuart, A. (1973). *The advanced theory of statistics, Vol.2: Inference and relationship.* 3rd edition. London: Griffin.

Kirkpatrick, S., Gellat, C.D. and Vecchi, M.P. (1983). Optimization by simulated annealing. *Science.* 220, pp.671-680.

Köhle, M. and Merkl, D. (1996). Identification of gait patterns with self-organizing maps based on ground reaction force. *Proc. of the European Symposium on Artificial Neural Networks (ESANN'96), Bruges, Belgium.*

Lafortune, M.A., Cavanagh, P.R., Sommer III, H. J. and Kalenak, A. (1992). Three-dimensional kinematics of the human knee during walking. *Journal of Biomechanics.* 25, 4, pp.347-357.

Lafuente R., Belda J.M., Sanchez-Lacuesta J., Soler C. and Prat J. (1998). Design and test of neural networks and statistical classifiers in computer-aided movement analysis: a case study on gait analysis. *Clinical Biomechanics.* 13, 3, pp.216-229.

Lipschitz, R. and Strauss, O. (1997). Coping with uncertainty: a naturalistic decision-making process. *Organizational behaviour and human decision processes*. 69, 2, pp.149-163.

Meglan, D. and Todd, F. (1994). "Kinetics of Human Locomotion" in Human Walking, 2nd edition, eds. Rose J. and Gamble J.G. Williams and Wilkins, Baltimore, Maryland.

Mellouli, K. and Elouedi, Z. (1997). Pooling expert opinions using Dempster-Shafer theory of evidence 'In: *Systems, Man and Cybernetics, IEEE International Conference on Computational Cybernetics and Simulation*'. 2, pp.1900-1905.

Merkle, L.A., Layne, C.S., Bloomberg, J.J. and Zhang, J.J. (1998). Using factor analysis to identify neuromuscular synergies during treadmill walking. *Journal of Neuroscience Methods*. 82, pp.207-214.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equation of state calculation by fast computing machines. *Journal of Chemical Physics*. 21, pp.1087-1091.

Moran, C.G. and Horton, T.C. (2000). Total knee replacement: the joint of the decade. *British Medical Journal*. 320, pp.820.

Myles, C.M., Rowe, P.J., Walker C.R.C. and Nutton R.W. (2002). Knee joint functional range of movement prior to and following total knee arthroplasty measured using flexible electrogoniometry. *Gait and Posture*. 16, pp.46-54.

Nevitt, M.C. and Lane, N. (1999). Body weight and osteoarthritis. *The American Journal of Medicine*. 107, pp.632-633.

Ng, K-C. and Abramson, B. (1990). Uncertainty management in expert systems. *IEEE Expert*. 5, 3, pp.29-48.

Nolle, L., Goodyear, A., Hopgood, A.A., Picton, P.D. and Braithwaite, N.St.J. (2002). Automated control of an actively compensated Langmuir probe system using simulated annealing. *Knowledge Based Systems*. 15, pp.349-354.

Olree, K.S. and Vaughan, C.L. (1995). Fundamental patterns of bilateral muscle activity in human locomotion. *Biological Cybernetics*. 73, pp.409-414.

O'Malley, M.J., Abel, M.F., Damiano, D.L. and Vaughan, C.L. (1997). Fuzzy clustering of children with cerebral palsy based on temporal-distance gait parameters. *IEEE Transactions on Rehabilitation Engineering*. 5, 4, pp.300-309.

Raudys, S.J., and Jain, A.K. (1991). Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 13, 3, pp.252-264.

Raymer, M.L., Punch, W.F., Goodman, E.D., Kuhn, L.A. and Jain, A.K. (2000). Dimensionality reduction using genetic algorithms. *IEEE Transactions on Evolutionary Computing*. 4, 2, pp.164-171.

Renner, G. and Ekárt, A. (2003). Genetic algorithms in computer aided design. *Computer-Aided Design*. 35, pp.709-726.

Rose J. and Gamble J.G. eds. (1994). "Kinetics of Human Locomotion" in Human Walking, 2nd edition, Williams and Wilkins, Baltimore, Maryland.

Roy, S., Wilson, C., Williams, R., Sharma, A.J., Holt, C., O'Callaghan, P. and Jones, L. (2002). Mobile bearing versus fixed bearing total knee replacement: A prospective randomised trial. *Proceedings of the British Association for Surgery of the Knee*, Spring meeting.

Sadeghi, H., Allard, P. and Duhaime, M. (1997). Functional gait asymmetry in able-bodied subjects. *Human Movement Science*. 16, pp.243-258.

Sadeghi, H., Prince, F., Sadeghi, S. and Labelle, H. (2000). Principal component analysis of the power developed in the flexion/extension muscles of the hip in able-bodied gait. *Medical Engineering & Physics*. 22, pp.703-710.

Safranek, R.J., Gottschlich, S. and Kak, A.C. (1990). Evidence accumulation using binary frames of discernment for verification vision, *IEEE Transactions on Robotics and Automation*, 6, 4, pp.405-417.

Schneider, E. and Chao, E.Y. (1983). Fourier analysis of ground reaction forces in normals and patients with knee joint disease. *Journal of Biomechanics*. 16, 8, pp.591-601.

Schutte, L.M., Narayanan, U., Stout, J.L., Selber, P., Gage, J.R. and Schwartz, M.H. (2000). An index for quantifying deviations from normal gait. *Gait and Posture*. 11, pp.25-31.

Shafer, G. (1976). *A Mathematical theory of Evidence*, Princeton: Princeton University Press.

Shafer, G. and Pearl, J. eds. (1990). *Readings in Uncertain Reasoning*. San Mateo, California: Kaufmann.

Shiavi, R. and Griffin, P. (1981). Representing and clustering electromyographic gait patterns with multivariate techniques. *Medical & Biological Engineering & Computing*. 19, pp.605-611.

Siedlecki, W. and Sklansky, J. (1988). On automatic feature selection. *International Journal of Pattern Recognition and Artificial Intelligence*. 2, 2, pp.197-220.

Siedlecki, W. and Sklansky, J. (1989). A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters*. 10, 5, pp.335-347.

Silvester P.P. and Ferrari R.L. (1996). *Finite Elements for Electrical Engineers*. 3rd edition. Cambride; New York: Cambridge University Press.

Smets, P. (1991). Varieties of ignorance and the need for well-founded theories. *Information Sciences*. 57-58, pp.135-144.

Söderkvist, I. and Wedin, P.A. (1993). Determining the movements of the skeleton using well configured markers, *Journal of Biomechanics*, 26, 12, pp.1473-1477.

Stroud, K. A. (1984). *Fourier series and harmonic analysis*. Cheltenham: Thornes.

Su, F-C. and Wu, W-L. (2000). Design and testing of a genetic algorithm neural network in the assessment of gait patterns, *Medical Engineering & Physics*. 22, pp.67-74.

Su, F-C., Wu, W-L., Cheng, Y-M. and Chou, Y-L. (2001). Fuzzy clustering of gait patterns of patients after ankle arthrodesis based on kinematic parameters. *Medical Engineering & Physics*. 23, pp.83-90.

Tabachnick, B.G. and Fidell, L.S. (1989). *Using multivariate statistics*. 2nd edition. Cambridge; Philadelphia: Harper & Row.

Toussaint, G.T. (1974). Bibliography on estimation of misclassification. *IEEE Transactions on Information Theory*. 20, 4, pp.472-479.

Walker, P.S. and Sathasivam, S. (2000). Design forms of total knee replacement. *Proc Instn Mech Engrs*. 214, H, pp.101-119.

Weiss, S.M. and Kulikowski, C.A. (1991). *Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems*. San Mateo, California: M. Kaufmann.

Whittle, M.W. and Jefferson, R.J. (1989). Functional biomechanical assessment of the Oxford meniscal knee. *Journal of Arthroplasty*. 4,3, pp.231-243.

Whittle, M.W. (1996). Gait analysis: an introduction. 2nd Edition. Oxford; Boston: Butterworth-Heinemann.

Wu, W-L. and Su, F-C. (2000). Potential of the back propagation neural network in the assessment of gait patterns in ankle arthrodesis. *Clinical Biomechanics*. 15, pp.143-145.

Yamamoto, S., Suto, Y., Kawamura, H., Hashizume, T., Kakurai, S. and Sugahara, S. (1983). Quantitative gait evaluation of hip diseases using principal component analysis. *Journal of Biomechanics*. 16, 9, pp.717-726.

Yoon, Y., Swales Jr., G. and Margavio, T.M. (1993). A comparison of discriminant analysis versus artificial neural networks. *The Journal of the Operational Research Society*. 44, 1, pp.51-60.

Zadeh L.A. (1965) *Inform. Control*. 8, pp.338-353.

Zatsiorsky, V.M. (1998). Kinematics of human motion. Champaign, IL: Human Kinetics.

APPENDIX A

KNEE OUTCOME SURVEY (Irrgang *et al.*, 1998)

SYMPTOMS:

1. **To what degree does pain in your knee affect your daily activity level?**

   ☐ I never have pain in my knee.
   ☐ I have pain in my knee, but it does not affect my daily activity.
   ☐ Pain affects my activity slightly.
   ☐ Pain affects my activity moderately.
   ☐ Pain affects my activity severely.
   ☐ Pain in my knee prevents me from performing all daily activities.

2. **To what degree does grinding or grating of your knee affect your daily activity level?**

   ☐ I never have grinding or grating in my knee.
   ☐ I have grinding or grating in my knee, but it does not affect my daily activity.
   ☐ Grinding or grating affects my activity slightly.
   ☐ Grinding or grating affects my activity moderately.
   ☐ Grinding or grating affects my activity severely.
   ☐ Grinding or grating in my knee prevents me from performing all daily activities.

3. **To what degree does stiffness in your knee affect your daily activity level?**

   ☐ I never have stiffness in my knee.
   ☐ I have stiffness in my knee, but it does not affect my daily activity.
   ☐ Stiffness affects my activity slightly.
   ☐ Stiffness affects my activity moderately.
   ☐ Stiffness affects my activity severely.
   ☐ Stiffness in my knee prevents me from performing all daily activities.

4. **To what degree does swelling in your knee affect your daily activity level?**

   ☐ I never have swelling in my knee.
   ☐ I have swelling in my knee, but it does not affect my daily activity.
   ☐ Swelling affects my activity slightly.
   ☐ Swelling affects my activity moderately.
   ☐ Swelling affects my activity severely.
   ☐ Swelling in my knee prevents me from performing all daily activities.

5. **To what degree does slipping of your knee affect your daily activity level?**

   ☐ I never have slipping of my knee.

☐ I have slipping in my knee, but it does not affect my daily activity.
☐ Slipping affects my activity slightly.
☐ Slipping affects my activity moderately.
☐ Slipping affects my activity severely.
☐ Slipping of my knee prevents me from performing all daily activities.

6. To what degree does buckling of your knee affect your daily activity level?

☐ I never have buckling of my knee.
☐ I have buckling of my knee, but it does not affect my daily activity.
☐ Buckling affects my activity slightly.
☐ Buckling affects my activity moderately.
☐ Buckling affects my activity severely.
☐ Buckling of my knee prevents me from performing all daily activities.

7. To what degree does weakness or lack of strength of your leg affect your daily activity level?

☐ My leg never feels weak.
☐ My leg feels weak, but it does not affect my daily activity.
☐ Weakness affects my activity slightly.
☐ Weakness affects my activity moderately.
☐ Weakness affects my activity severely.
☐ Weakness of my leg prevents me from performing all daily activities.

FUNCTIONAL DISABILITY WITH ACTIVITIES OF DAILY LIVING:

8. How does your knee affect your ability to walk?

☐ My knee does not affect my ability to walk.
☐ I have pain in my knee when walking, but it does not affect my ability to walk.
☐ My knee prevents me from walking more than 1 mile.
☐ My knee prevents me from walking more than 1/2 mile.
☐ My knee prevents me from walking more than 1 block.
☐ My knee prevents me from walking.

9. Because of your knee, do you walk with crutches or a cane?

☐ I can walk without crutches or a cane.
☐ My knee causes me to walk with 1 crutch or a cane.
☐ My knee causes me to walk with 2 crutches.
☐ Because of my knee, I cannot walk even with crutches.

10. Does your knee cause you to limp when you walk?

☐ I can walk without a limp.
☐ Sometimes my knee causes me to walk with a limp.
☐ Because of my knee, I cannot walk without a limp.

11. How does your knee affect your ability to go up stairs?

- ☐ My knee does not affect my ability to go up stairs.
- ☐ I have pain in my knee when going up stairs, but it does not limit my ability to go up stairs.
- ☐ I am able to go up stairs normally, but I need to rely on use of a railing.
- ☐ I am able to go up stairs one step at a time with use of a railing.
- ☐ I have to use crutches or a cane to go up stairs.
- ☐ I cannot go up stairs.

12. How does your knee affect your ability to go down stairs?

- ☐ My knee does not affect my ability to go down stairs.
- ☐ I have pain in my knee when going down stairs, but it does not limit my ability to go down stairs.
- ☐ I am able to go down stairs normally, but I need to rely on use of a railing.
- ☐ I am able to go down stairs one step at a time with use of a railing.
- ☐ I have to use crutches or a cane to go down stairs.
- ☐ I cannot go down stairs.

13. How does your knee affect your ability to stand?

- ☐ My knee does not affect my ability to stand, I can stand for unlimited amounts of time.
- ☐ I have pain in my knee when standing, but it does not limit my ability to stand.
- ☐ Because of my knee I cannot stand for more than 1 hour.
- ☐ Because of my knee I cannot stand for more than 1/2 hour.
- ☐ Because of my knee I cannot stand for more than 10 minutes.
- ☐ I cannot stand because of my knee.

14. How does your knee affect your ability to kneel on the front of your knee?

- ☐ My knee does not affect my ability to kneel on the front of my knee. I can kneel for unlimited amounts of time.
- ☐ I have pain when kneeling on the front of my knee, but it does not limit my ability to kneel.
- ☐ I cannot kneel on the front of your knee for more than 1 hour.
- ☐ I cannot kneel on the front of your knee for more than 1/2 hour.
- ☐ I cannot kneel on the front of your knee for more than 10 minutes.
- ☐ I cannot kneel on the front of your knee.

15. How does your knee affect your ability to squat?

- ☐ My knee does not affect my ability to squat, I can squat all the way down.
- ☐ I have pain in my knee when squatting, but I can still squat all the way down.
- ☐ I cannot squat more than3/4 of the way down.
- ☐ I cannot squat more than 1/2 of the way down.

☐ I cannot squat more than 1/4 of the way down.
☐ I cannot squat because of my knee.

16. How does your knee affect your ability to sit with your knee bent?

☐ My knee does not affect my ability to sit with my knee bent, I can sit for unlimited amounts of time.
☐ I have pain in my knee when sitting with my knee bent, but it does not limit my ability to sit.
☐ I cannot sit with my knee bent for more than 1 hour.
☐ I cannot sit with my knee bent for more than 1/2 hour.
☐ I cannot sit with my knee bent for more than 10 minutes.
☐ I cannot sit with my knee bent.

17. How does your knee affect your ability to rise from a chair?

☐ My knee does not affect my ability to rise from a chair.
☐ I have pain when rising from a seated position, but it does not affect my ability to rise from a seated position.
☐ Because of my knee I can only rise from a chair if I use my hands and arms to assist.
☐ Because of my knee I cannot rise from a chair.

APPENDIX B

Table B.1 The sensitivity of k (k_c and k_s) to changes in population for each input variable, v_i (i = 1:18)

Input Variable	k_c		k_s	
	Mean	**Standard Deviation**	**Mean**	**Standard Deviation**
1	0.1612	0.0328	0.1307	0.0054
2	-0.7468	0.0087	-0.1425	0.0024
3	0.5561	0.0161	0.5231	0.0109
4	-0.7636	0.0081	-0.1362	0.0018
5	0.1000	0.0259	0.2055	0.0077
6	0.3349	0.0197	0.3504	0.0080
7	-0.0222	0.0407	-0.1250	0.0927
8	0.8473	0.0060	0.1709	0.0019
9	-0.0052	0.0304	-0.0409	0.2789
10	0.0234	0.0265	0.1001	0.0878
11	0.7734	0.0079	0.2064	0.0034
12	0.2538	0.0248	0.1292	0.0030
13	0.0864	0.0258	0.2333	0.0050
14	-0.3820	0.0251	-0.3093	0.0074
15	0.3048	0.0216	0.1287	0.0024
16	0.4934	0.0246	0.6909	0.0127
17	0.4547	0.0205	0.9145	0.0151
18	0.1812	0.0281	0.2004	0.0049

Table B.2 Test 1 accuracy, % and OB results

Subtest	Part	Accuracy, %		OB	
		In-sample	Out-of-sample	In-sample	Out-of-sample
I	a	97.56	97.62	0.0368	0.0432
	b	97.62	97.62	0.0370	0.0455
	c	97.62	97.62	0.0424	0.0556
	d	97.62	97.62	0.0529	0.0707
	e	97.62	97.62	0.0698	0.0914
	f	97.62	97.62	0.0965	0.1207
	g	97.62	97.62	0.1398	0.1647
	h	97.62	97.62	0.2119	0.2350
	i	97.62	97.62	0.3363	0.3541
	j	97.62	97.62	0.5605	0.5692
II	a	92.80	83.33	0.0963	0.1760
	b	94.25	88.10	0.0995	0.1623
	c	94.25	88.10	0.1116	0.1713
	d	94.14	88.10	0.1296	0.1861
	e	94.37	88.10	0.1546	0.2077
	f	94.48	85.71	0.1895	0.2384
	g	94.72	88.10	0.2397	0.2827
	h	94.60	88.10	0.3147	0.3497
	i	94.60	88.10	0.4328	0.4568
	j	94.60	88.10	0.6308	0.6414

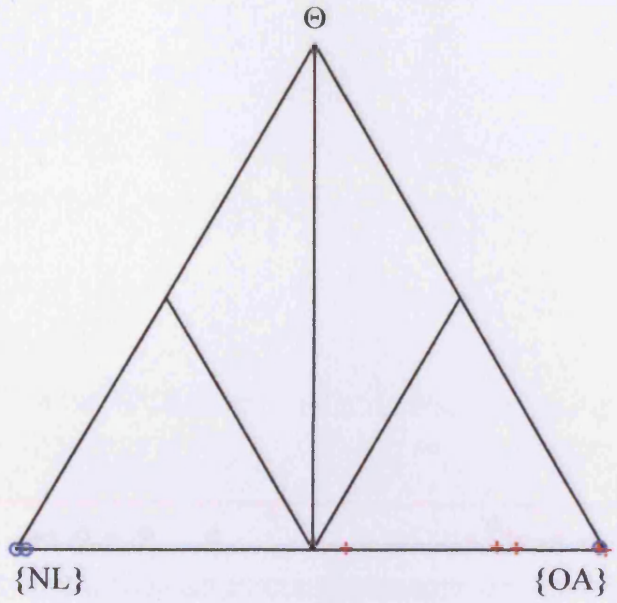


Figure B.1 Simplex plot showing simplex coordinates of out-of-sample subjects from test 1Ia ($[\Theta_L, \Theta_U] = [0, 1]$ and $k = k_c$)

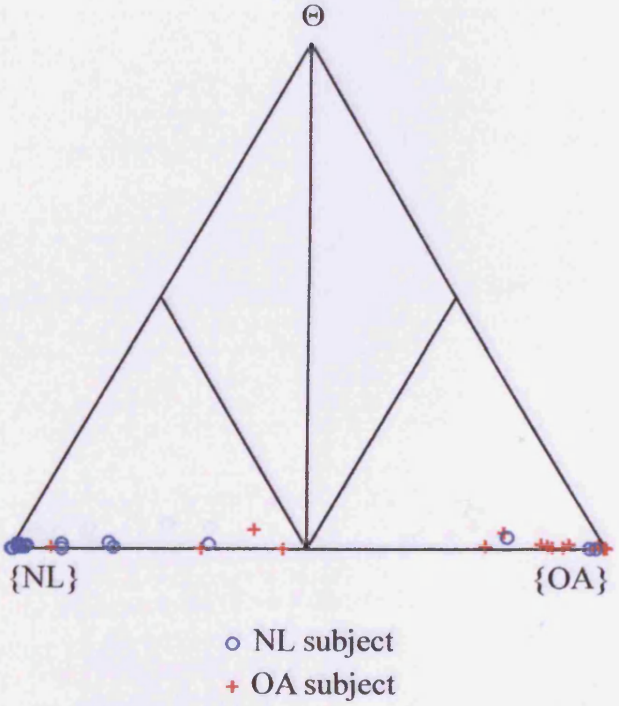


Figure B.2 Simplex plot showing simplex coordinates of out-of-sample subjects from test 1IIa ($[\Theta_L, \Theta_U] = [0,1]$ and $k = k_s$)

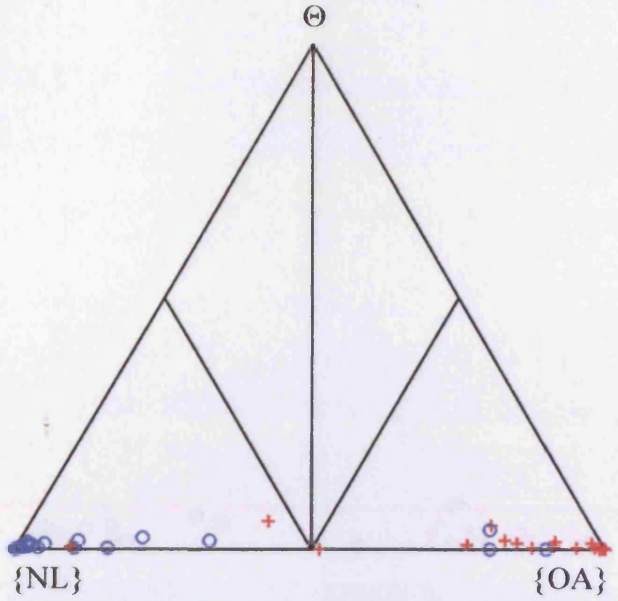


Figure B.3 Simplex plot showing simplex coordinates of out-of-sample subjects from test 1IIb ($[\Theta_L, \Theta_U] = [0.1, 1]$ and $k = k_s$)

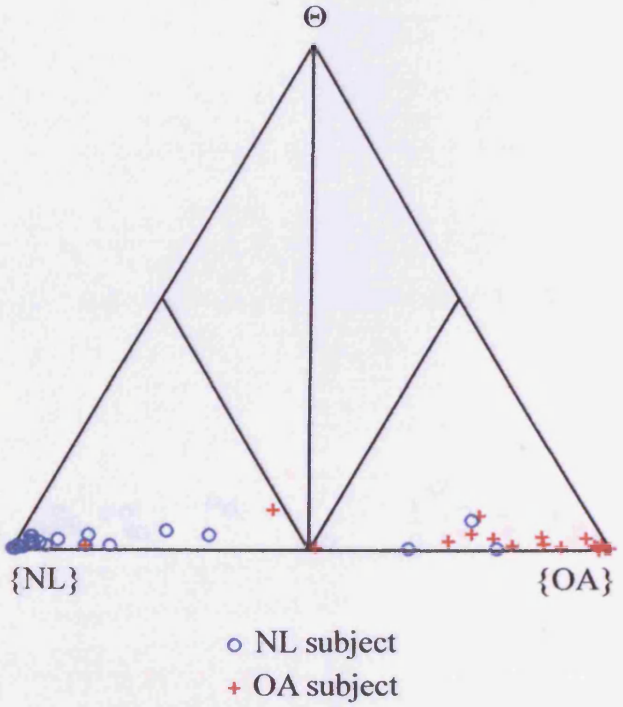


Figure B.4 Simplex plot showing simplex coordinates of out-of-sample subjects from test 1IIc ($[\Theta_L, \Theta_U] = [0.2, 1]$ and $k = k_s$)

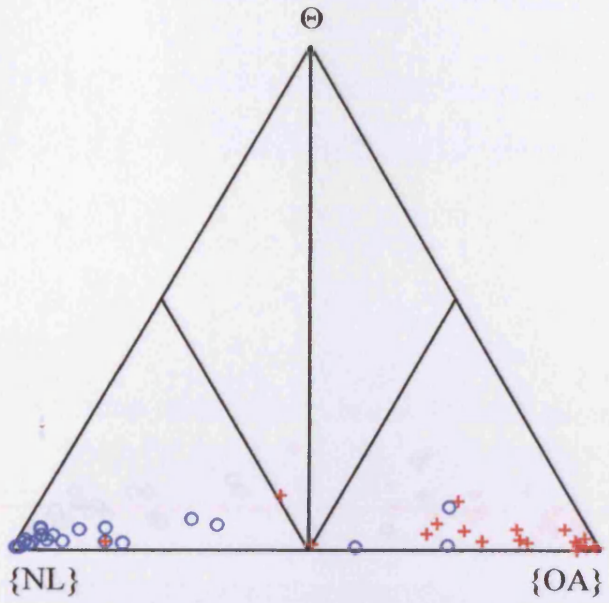


Figure B.5 Simplex plot showing simplex coordinates of out-of-sample subjects from test 1IId ($[\Theta_L, \Theta_U] = [0.3, 1]$ and $k = k_s$)

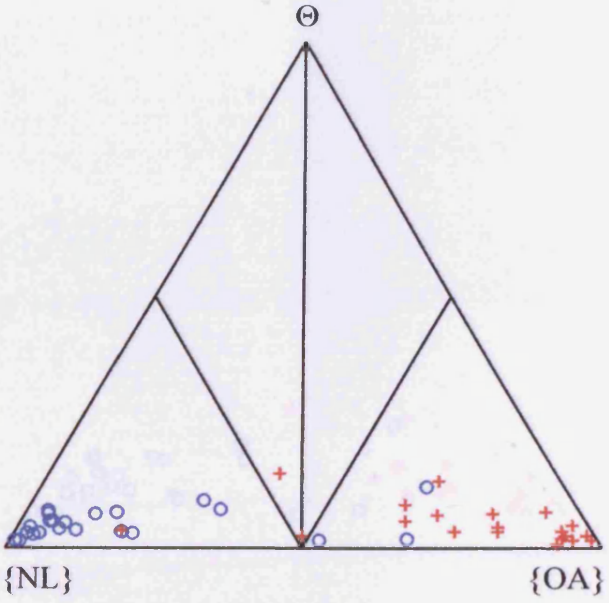


Figure B.6 Simplex plot showing simplex coordinates of out-of-sample subjects from test 1IIe ($[\Theta_L, \Theta_U] = [0.4, 1]$ and $k = k_s$)

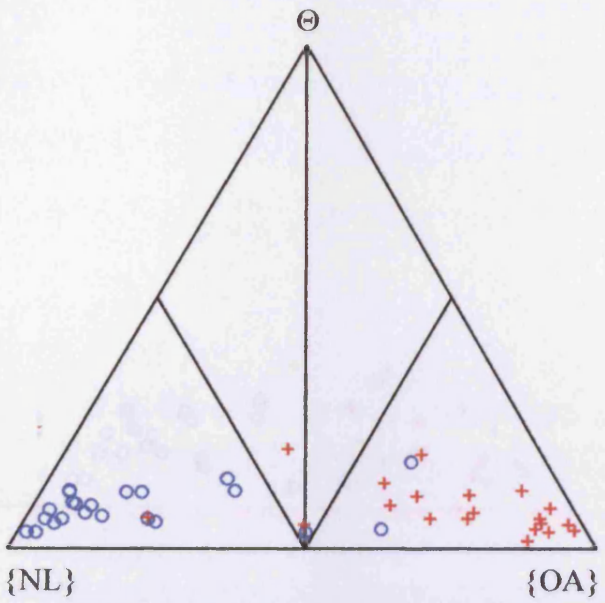


Figure B.7 Simplex plot showing simplex coordinates of out-of-sample subjects from test 1IIf ($[\Theta_L, \Theta_U] = [0.5, 1]$ and $k = k_s$)

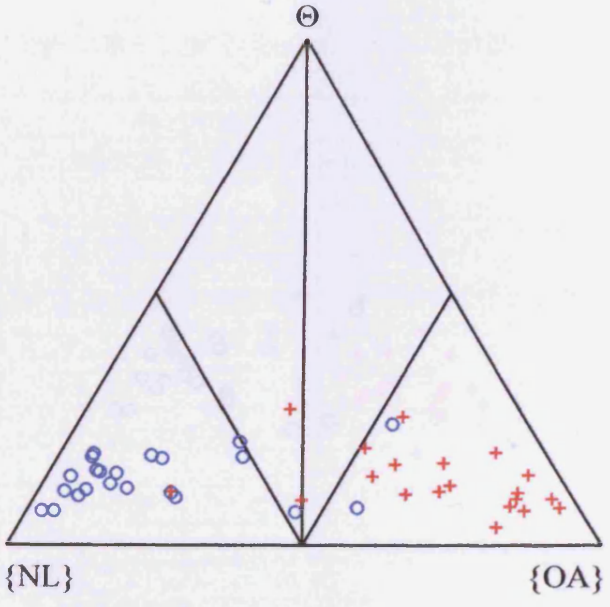


Figure B.8 Simplex plot showing simplex coordinates of out-of-sample subjects from test 1IIg ($[\Theta_L, \Theta_U] = [0.6, 1]$ and $k = k_s$)

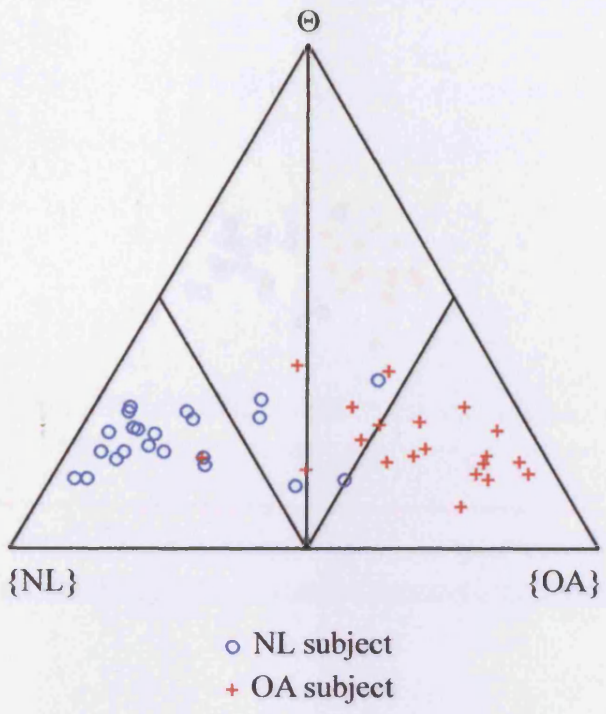


Figure B.9 Simplex plot showing simplex coordinates of out-of-sample subjects from test 1IIh ($[\Theta_L, \Theta_U] = [0.7, 1]$ and $k = k_s$)

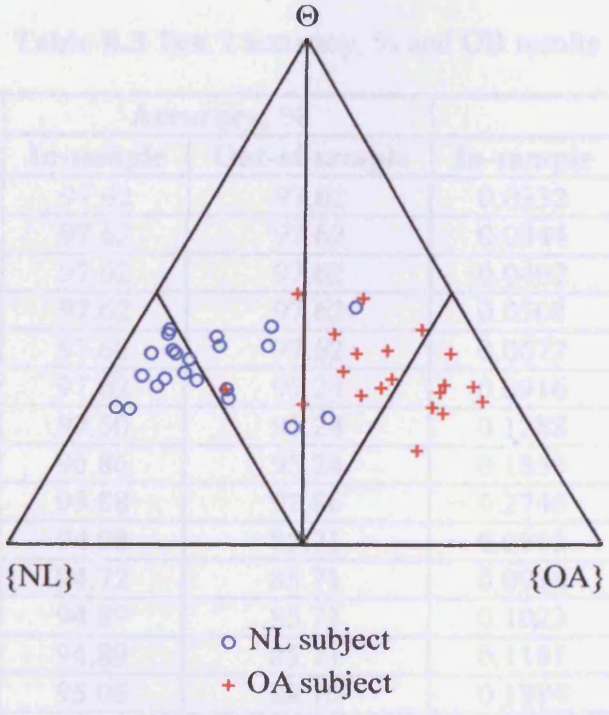


Figure B.10 Simplex plot showing simplex coordinates of out-of-sample subjects from test 1IIi ($[\Theta_L, \Theta_U] = [0.8, 1]$ and $k = k_s$)

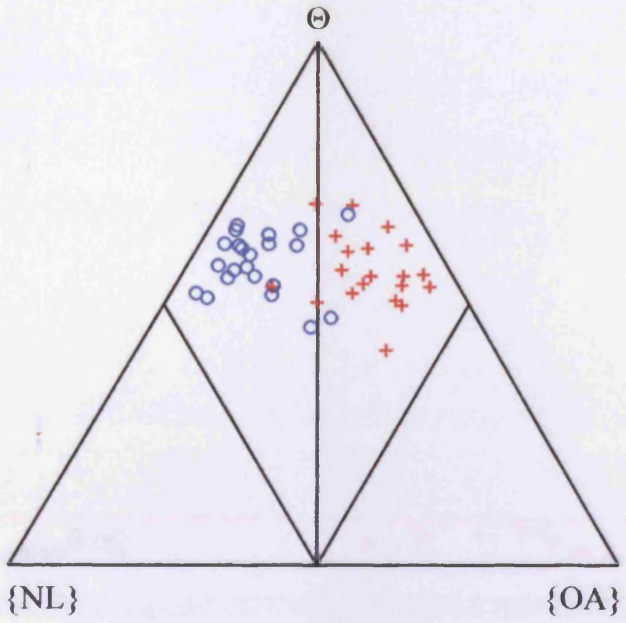


Figure B.11 Simplex plot showing simplex coordinates of out-of-sample subjects from test 1IIj ($[\Theta_L, \Theta_U] = [0.9, 1]$ and $k = k_s$)

Table B.3 Test 2 accuracy, % and OB results

Subtest	Part	Accuracy, %		OB	
		In-sample	Out-of-sample	In-sample	Out-of-sample
I	a	97.62	97.62	0.0332	0.0400
	b	97.62	97.62	0.0344	0.0432
	c	97.62	97.62	0.0402	0.0546
	d	97.62	97.62	0.0508	0.0702
	e	97.62	97.62	0.0672	0.0904
	f	97.62	95.24	0.0916	0.1177
	g	97.50	95.24	0.1288	0.1564
	h	96.86	95.24	0.1859	0.2129
	i	95.88	92.86	0.2746	0.2977
II	a	94.08	85.71	0.0885	0.1689
	b	94.72	85.71	0.0914	0.1551
	c	94.89	85.71	0.1023	0.1626
	d	94.89	85.71	0.1181	0.1755
	e	95.06	88.10	0.1394	0.1942
	f	95.06	88.10	0.1679	0.2197
	g	95.12	88.10	0.2064	0.2542
	h	94.89	90.48	0.2594	0.3012
	i	94.72	88.10	0.3336	0.3662

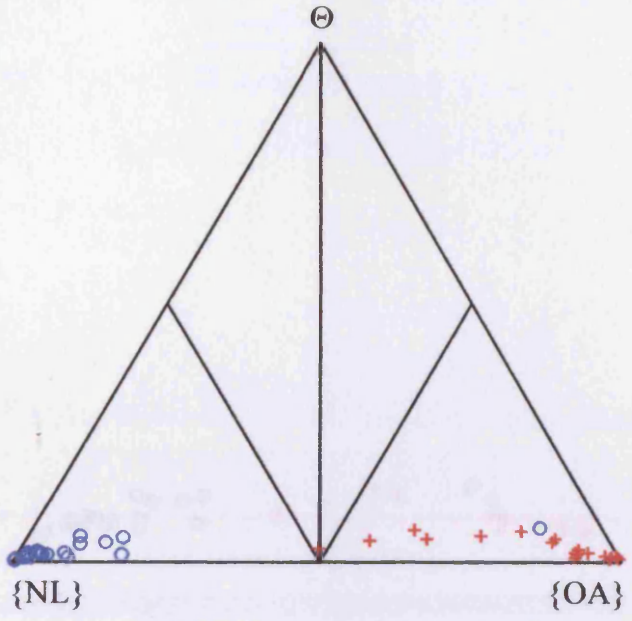


Figure B.12 Simplex plot showing simplex coordinates of out-of-sample subjects from test 2If ($[\Theta_L, \Theta_U] = [0.5, 0.9]$ and $k = k_c$)

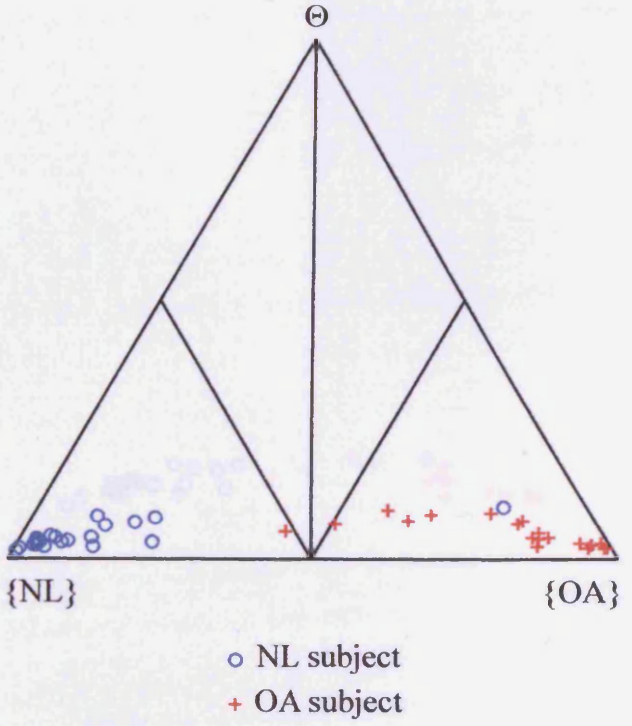


Figure B.13 Simplex plot showing simplex coordinates of out-of-sample subjects from test 2Ig ($[\Theta_L, \Theta_U] = [0.6, 0.9]$ and $k = k_c$)

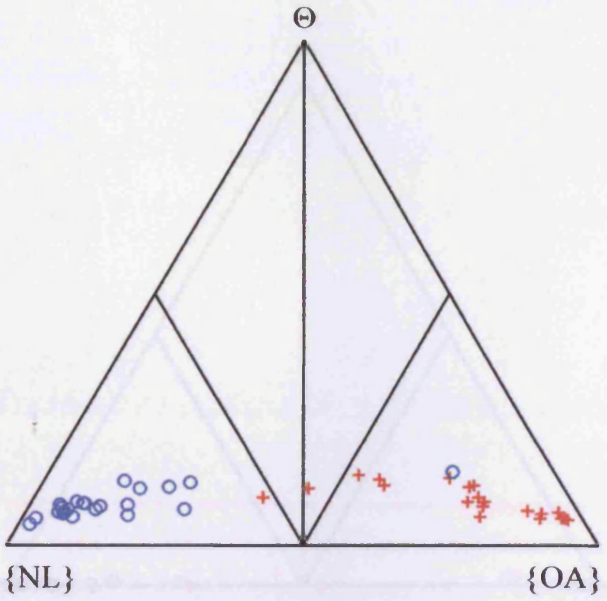


Figure B.14 Simplex plot showing simplex coordinates of out-of-sample subjects from test 2Ih ($[\Theta_L, \Theta_U] = [0.7, 0.9]$ and $k = k_c$)

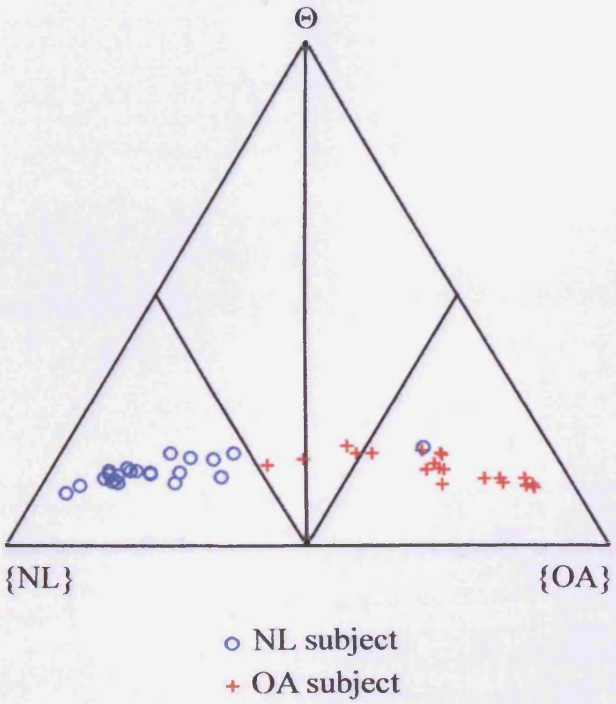


Figure B.15 Simplex plot showing simplex coordinates of out-of-sample subjects from test 2Ii ($[\Theta_L, \Theta_U] = [0.8, 0.9]$ and $k = k_c$)

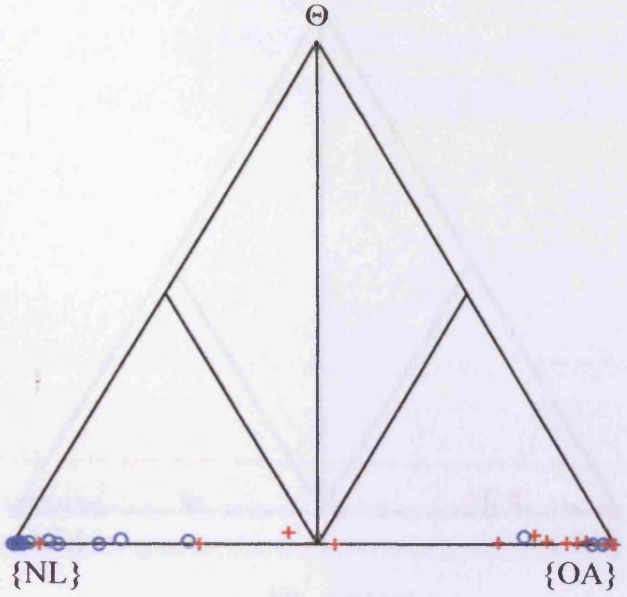


Figure B.16 Simplex plot showing simplex coordinates of out-of-sample subjects from test 2IIa ($[\Theta_L, \Theta_U] = [0, 0.9]$ and $k = k_s$)

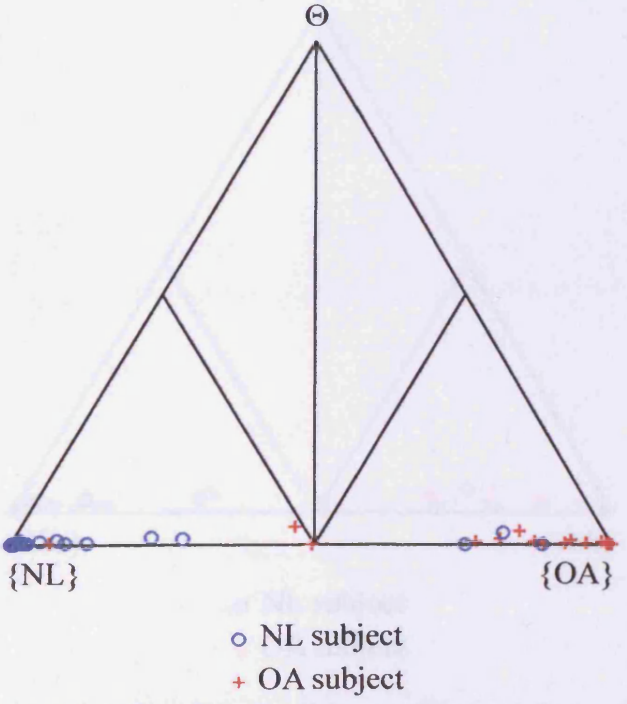
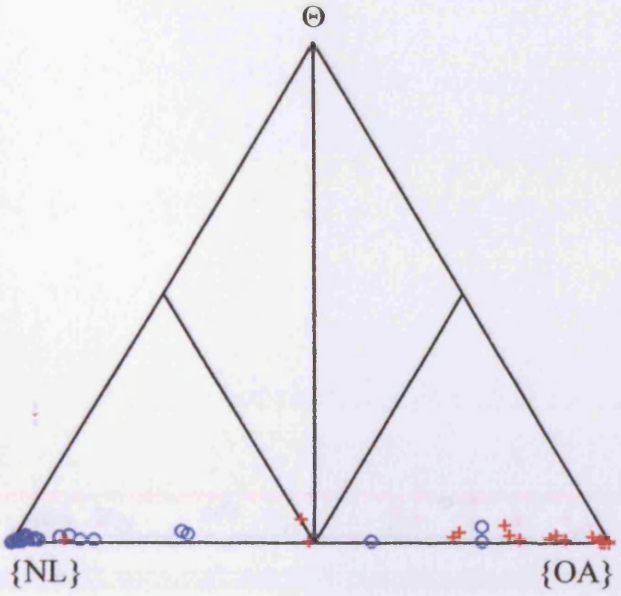
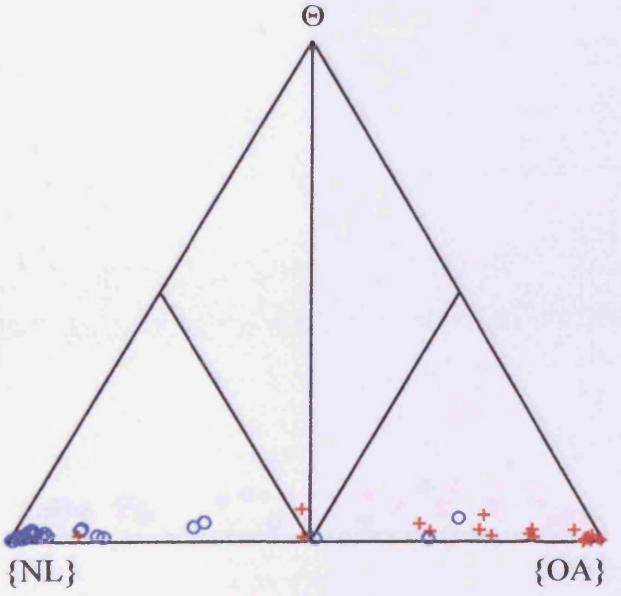


Figure B.17 Simplex plot showing simplex coordinates of out-of-sample subjects from test 2IIb ($[\Theta_L, \Theta_U] = [0.1, 0.9]$ and $k = k_s$)



o NL subject
+ OA subject

Figure B.18 Simplex plot showing simplex coordinates of out-of-sample subjects from test 2IIc ($[\Theta_L, \Theta_U] = [0.2, 0.9]$ and $k = k_s$)



o NL subject
+ OA subject

Figure B.19 Simplex plot showing simplex coordinates of out-of-sample subjects from test 2IId ($[\Theta_L, \Theta_U] = [0.3, 0.9]$ and $k = k_s$)

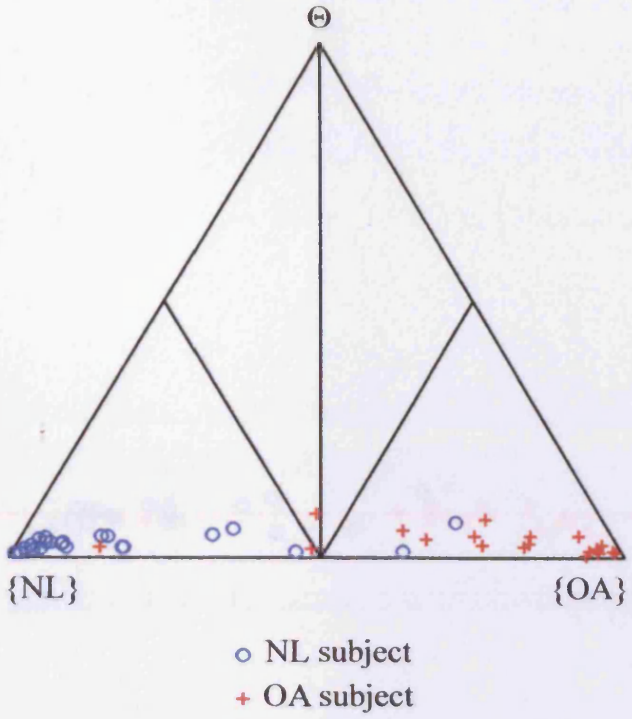


Figure B.20 Simplex plot showing simplex coordinates of out-of-sample subjects from test 2IIe ($[\Theta_L, \Theta_U] = [0.4, 0.9]$ and $k = k_s$)

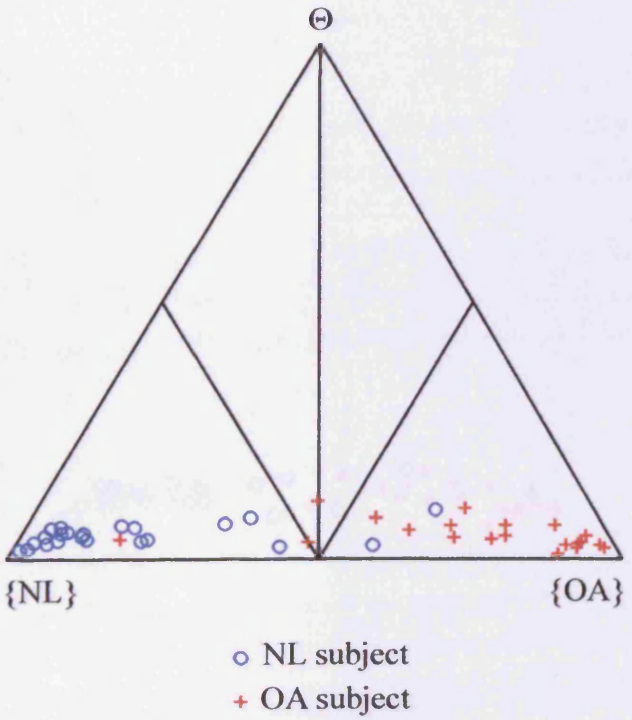


Figure B.21 Simplex plot showing simplex coordinates of out-of-sample subjects from test 2IIf ($[\Theta_L, \Theta_U] = [0.5, 0.9]$ and $k = k_s$)

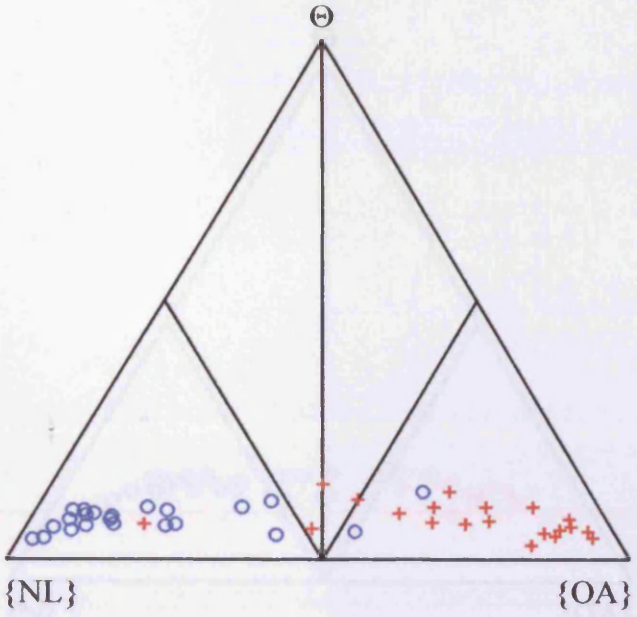


Figure B.22 Simplex plot showing simplex coordinates of out-of-sample subjects from test 2IIg ($[\Theta_L, \Theta_U] = [0.6, 0.9]$ and $k = k_s$)

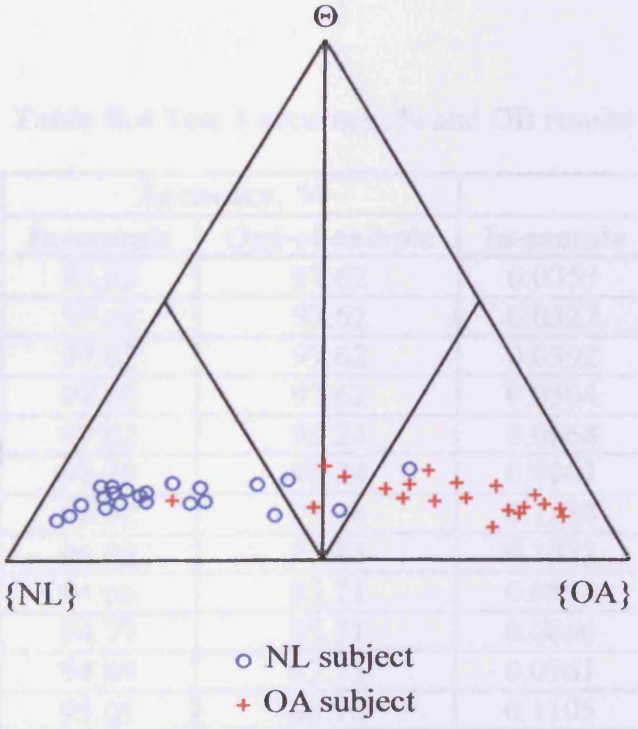


Figure B.23 Simplex plot showing simplex coordinates of out-of-sample subjects from test 2IIh ($[\Theta_L, \Theta_U] = [0.7, 0.9]$ and $k = k_s$)

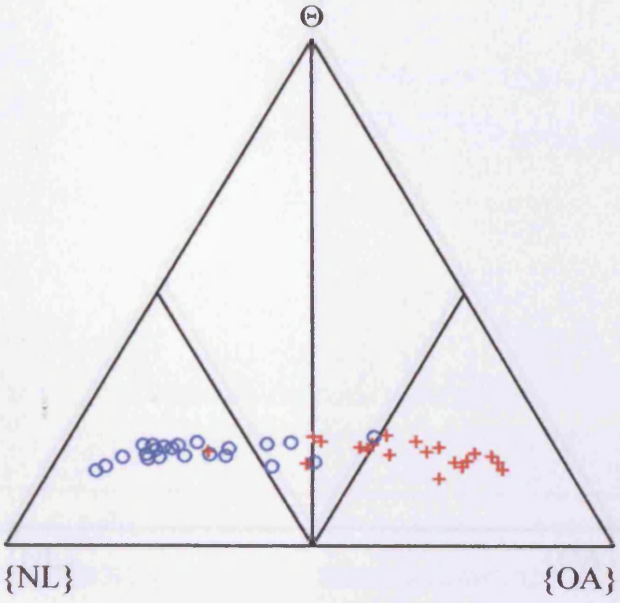


Figure B.24 Simplex plot showing simplex coordinates of out-of-sample subjects from test 2IIi ($[\Theta_L, \Theta_U] = [0.8, 0.9]$ and $k = k_s$)

Table B.4 Test 3 accuracy, % and OB results

Subtest	Part	Accuracy, %		OB	
		In-sample	Out-of-sample	In-sample	Out-of-sample
I	a	97.62	97.62	0.0353	0.0374
	b	97.62	97.62	0.0327	0.0419
	c	97.62	97.62	0.0392	0.0550
	d	97.62	97.62	0.0504	0.0714
	e	97.62	95.24	0.0668	0.0914
	f	97.39	95.24	0.0903	0.1176
	g	96.92	95.24	0.1258	0.1546
	h	96.98	95.24	0.1823	0.2104
II	a	94.66	85.71	0.0832	0.1637
	b	94.77	85.71	0.0860	0.1497
	c	94.89	85.71	0.0961	0.1559
	d	95.01	88.10	0.1105	0.1675
	e	95.06	88.10	0.1297	0.1843
	f	94.83	90.48	0.1549	0.2067
	g	94.83	88.10	0.1890	0.2371
	h	94.60	88.10	0.2419	0.2839

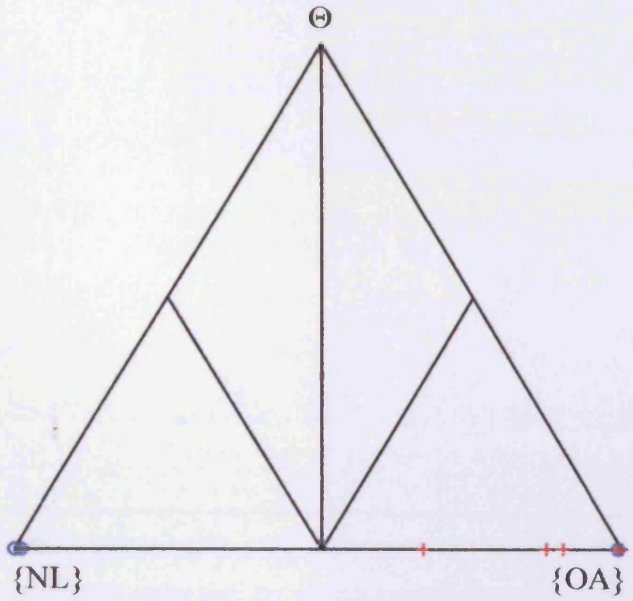


Figure B.25 Simplex plot showing simplex coordinates of out-of-sample subjects from test 3Ia ($[\Theta_L, \Theta_U] = [0, 0.8]$ and $k = k_c$)

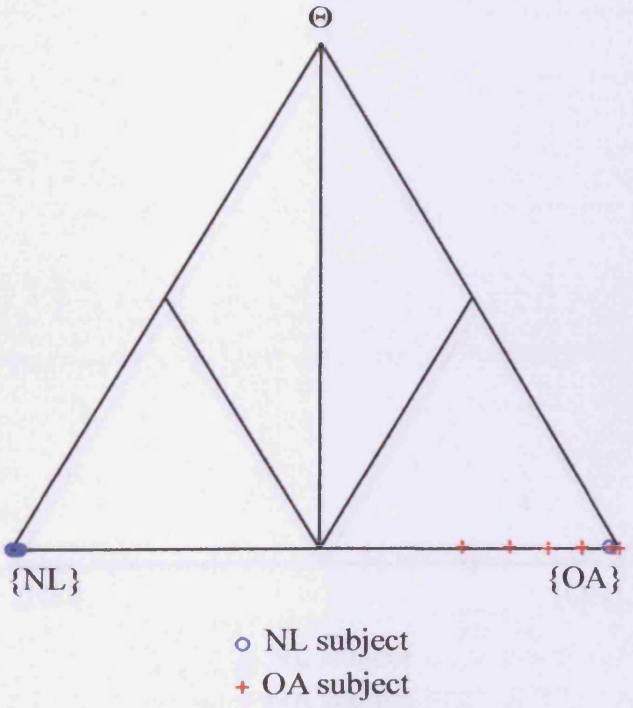


Figure B.26 Simplex plot showing simplex coordinates of out-of-sample subjects from test 3Ib ($[\Theta_L, \Theta_U] = [0.1\ 0.8]$ and $k = k_c$)

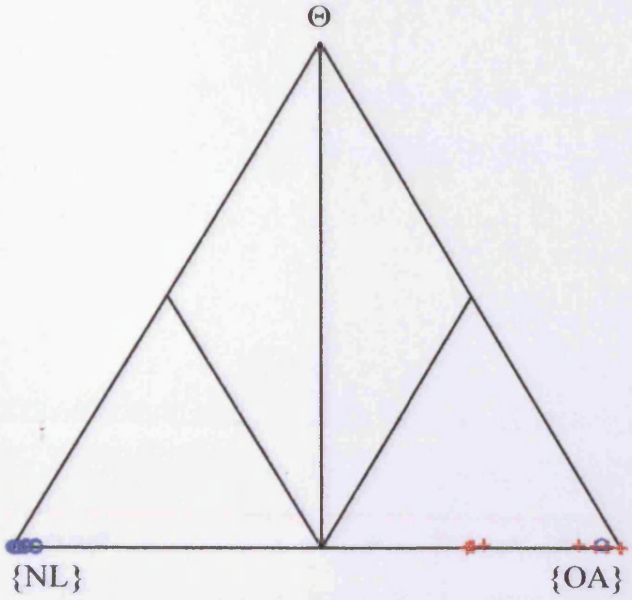


Figure B.27 Simplex plot showing simplex coordinates of out-of-sample subjects from test 3Ic ($[\Theta_L, \Theta_U] = [0.2, 0.8]$ and $k = k_c$)

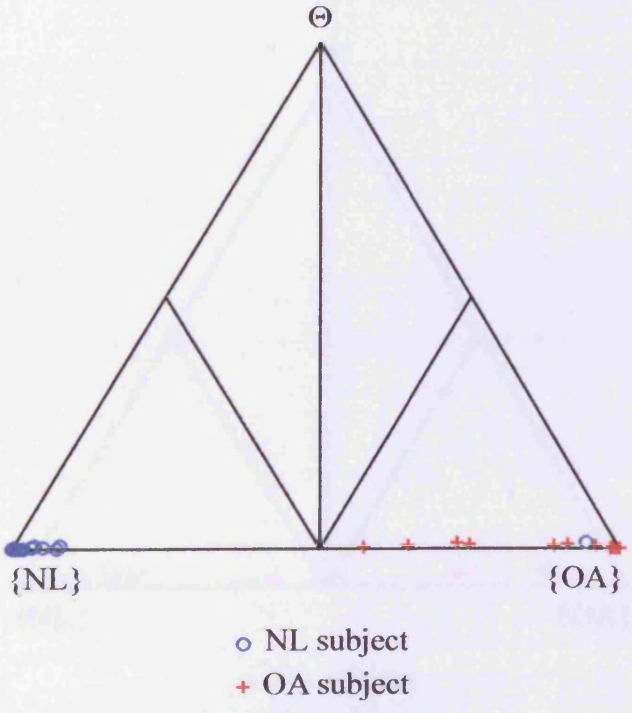


Figure B.28 Simplex plot showing simplex coordinates of out-of-sample subjects from test 3Id ($[\Theta_L, \Theta_U] = [0.3, 0.8]$ and $k = k_c$).

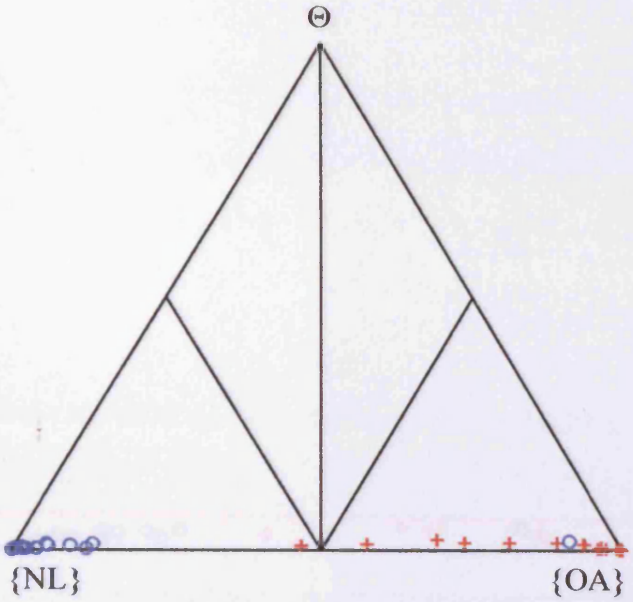


Figure B.29 Simplex plot showing simplex coordinates of out-of-sample subjects from test 3Ie ($[\Theta_L, \Theta_U] = [0.4, 0.8]$ and $k = k_c$)

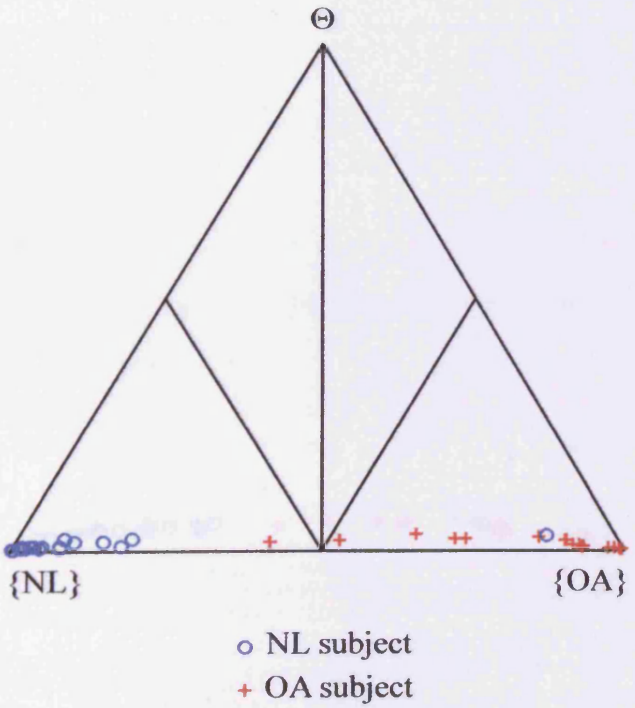


Figure B.30 Simplex plot showing simplex coordinates of out-of-sample subjects from test 3If ($[\Theta_L, \Theta_U] = [0.5, 0.8]$ and $k = k_c$)

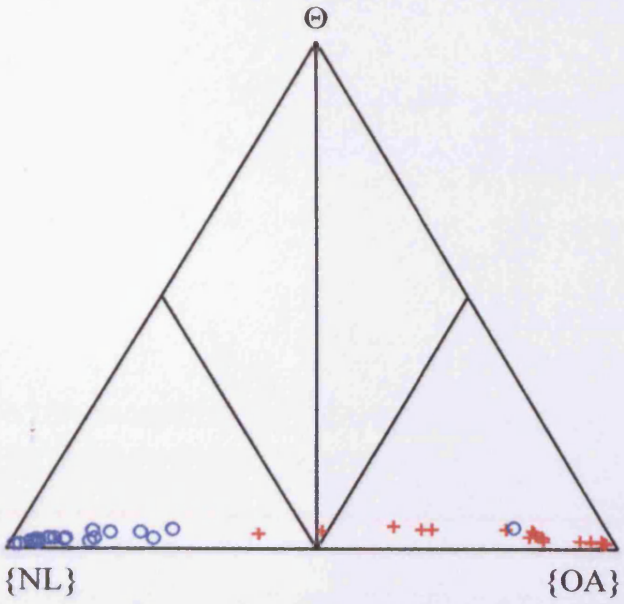


Figure B.31 Simplex plot showing simplex coordinates of out-of-sample subjects from test 3Ig ($[\Theta_L, \Theta_U] = [0.6, 0.8]$ and $k = k_c$)

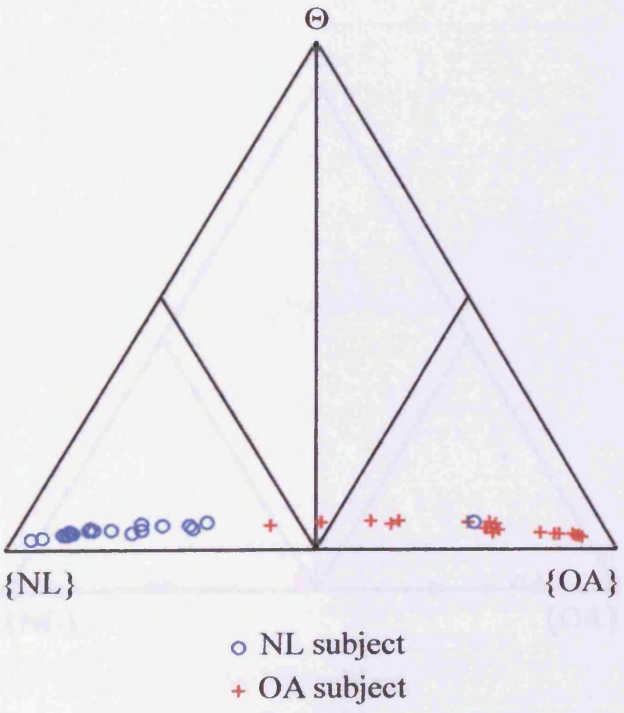


Figure B.32 Simplex plot showing simplex coordinates of out-of-sample subjects from test 3Ih ($[\Theta_L, \Theta_U] = [0.7, 0.8]$ and $k = k_c$)

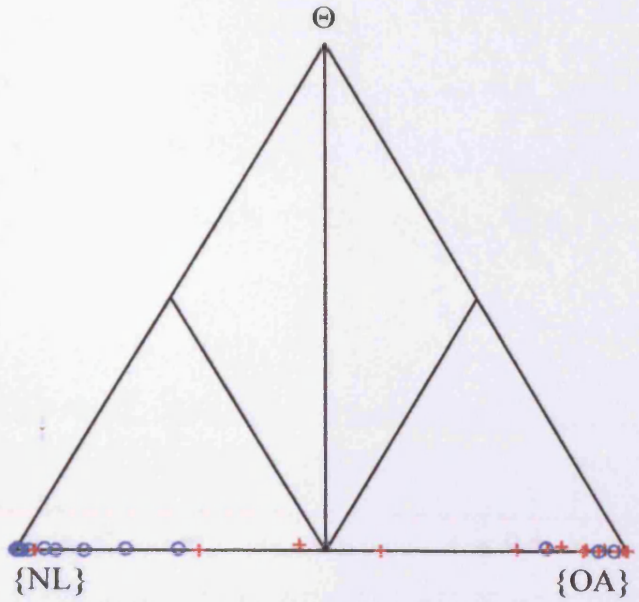


Figure B.33 Simplex plot showing simplex coordinates of out-of-sample subjects from test 3IIa ($[\Theta_L, \Theta_U] = [0, 0.8]$ and $k = k_s$)

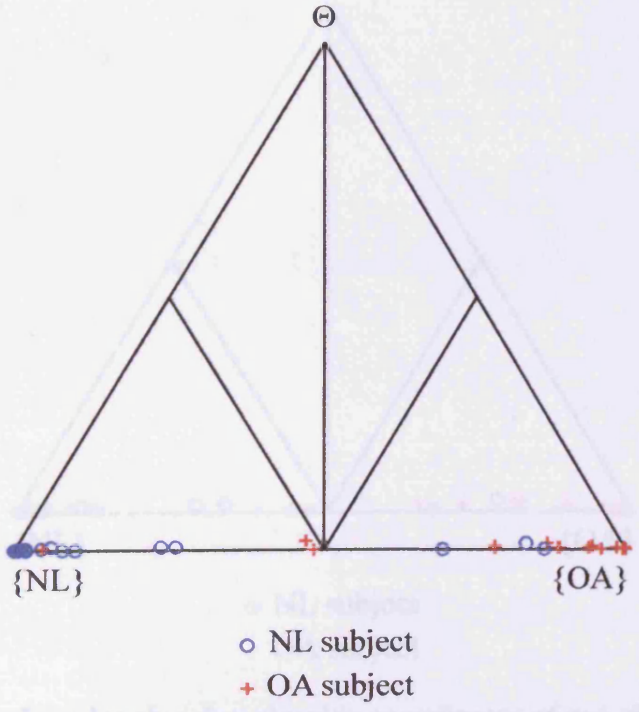


Figure B.34 Simplex plot showing simplex coordinates of out-of-sample subjects from test 3IIb ($[\Theta_L, \Theta_U] = [0.1, 0.8]$ and $k = k_s$)

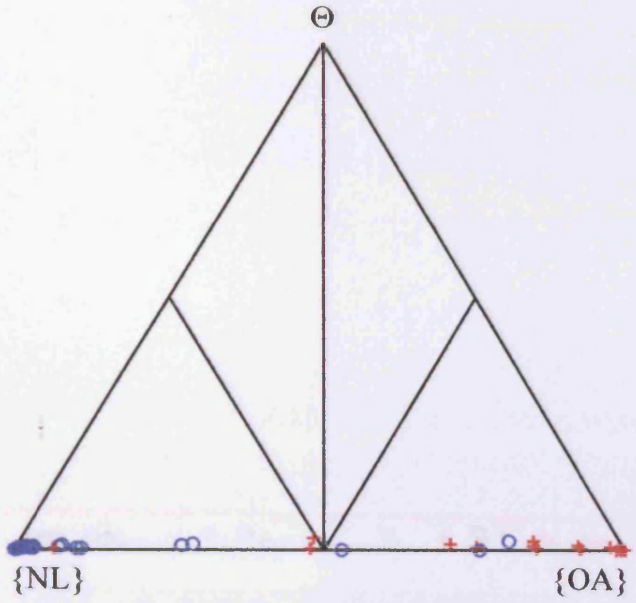


Figure B.35 Simplex plot showing simplex coordinates of out-of-sample subjects from test 3IIc ($[\Theta_L, \Theta_U] = [0.2, 0.8]$ and $k = k_s$)

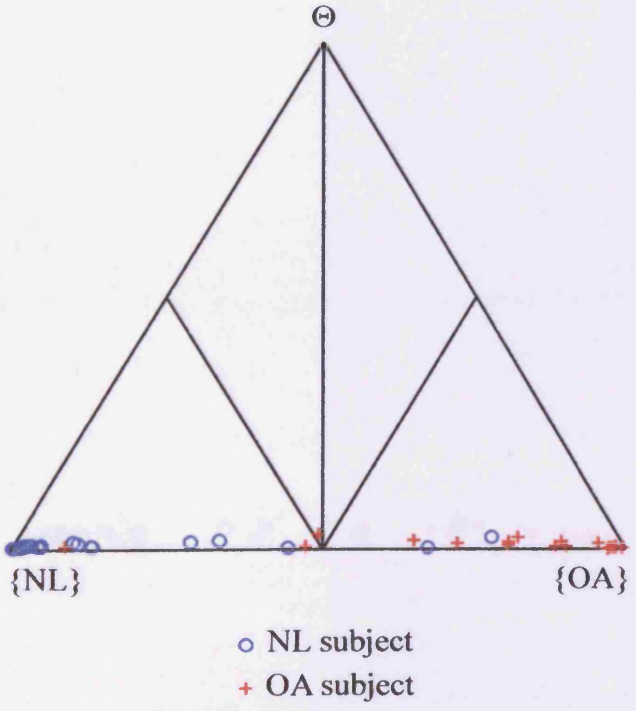


Figure B.36 Simplex plot showing simplex coordinates of out-of-sample subjects from test 3IId ($[\Theta_L, \Theta_U] = [0.3, 0.8]$ and $k = k_s$)

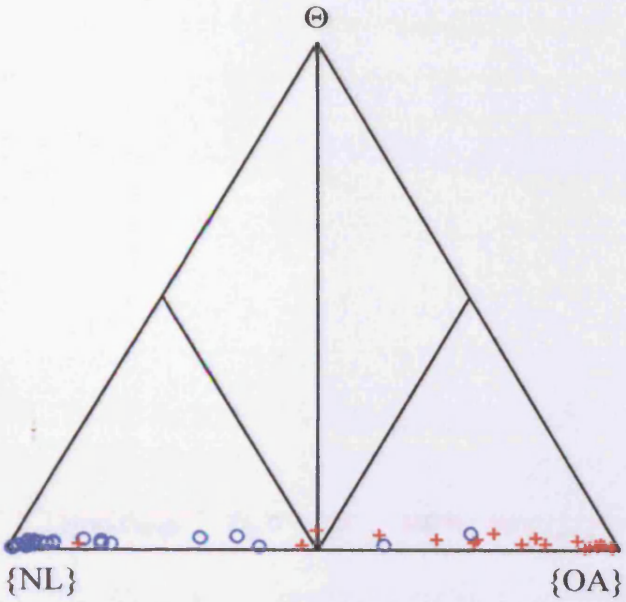


Figure B.37 Simplex plot showing simplex coordinates of out-of-sample subjects from test 3IIe ($[\Theta_L, \Theta_U] = [0.4, 0.8]$ and $k = k_s$)

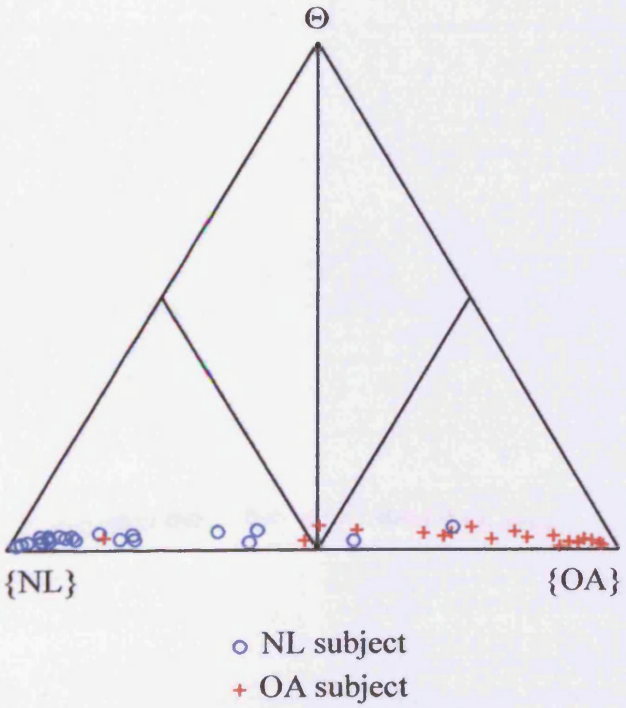


Figure B.38 Simplex plot showing simplex coordinates of out-of-sample subjects from test 3IIf ($[\Theta_L, \Theta_U] = [0.5, 0.8]$ and $k = k_s$)

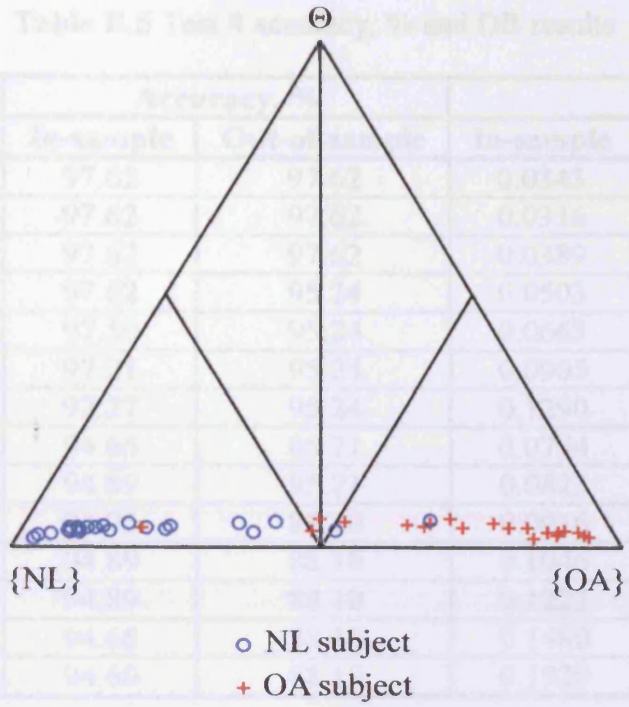


Figure B.39 Simplex plot showing simplex coordinates of out-of-sample subjects from test 3IIg ($[\Theta_L, \Theta_U] = [0.6, 0.8]$ and $k = k_s$)

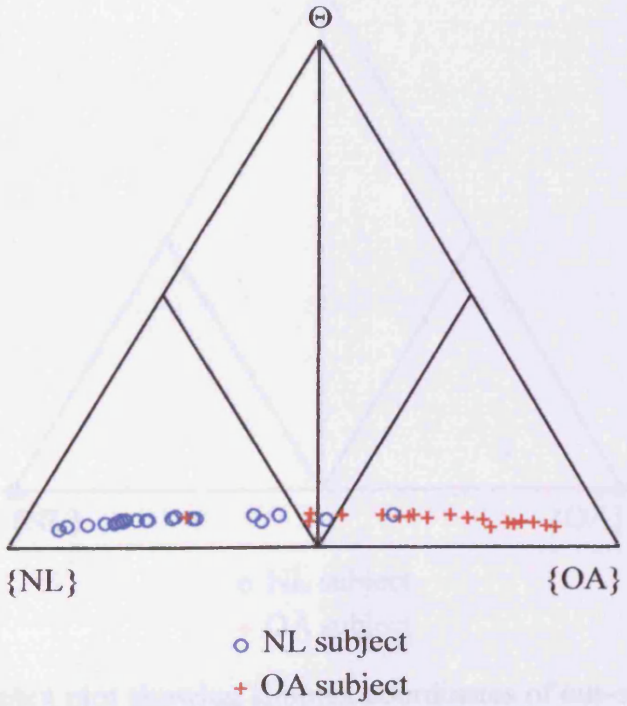


Figure B.40 Simplex plot showing simplex coordinates of out-of-sample subjects from test 3IIh ($[\Theta_L, \Theta_U] = [0.7, 0.8]$ and $k = k_s$)

Table B.5 Test 4 accuracy, % and OB results

Subtest	Part	Accuracy, %		OB	
		In-sample	Out-of-sample	In-sample	Out-of-sample
I	a	97.62	97.62	0.0343	0.0351
	b	97.62	97.62	0.0316	0.0417
	c	97.62	97.62	0.0389	0.0566
	d	97.62	95.24	0.0503	0.0727
	e	97.50	95.24	0.0663	0.0917
	f	97.21	95.24	0.0905	0.1184
	g	97.27	95.24	0.1290	0.1576
II	a	94.66	85.71	0.0794	0.1591
	b	94.89	85.71	0.0823	0.1447
	c	94.95	88.10	0.0916	0.1497
	d	94.89	88.10	0.1046	0.1600
	e	94.89	88.10	0.1223	0.1757
	f	94.66	88.10	0.1480	0.1991
	g	94.60	88.10	0.1920	0.2392

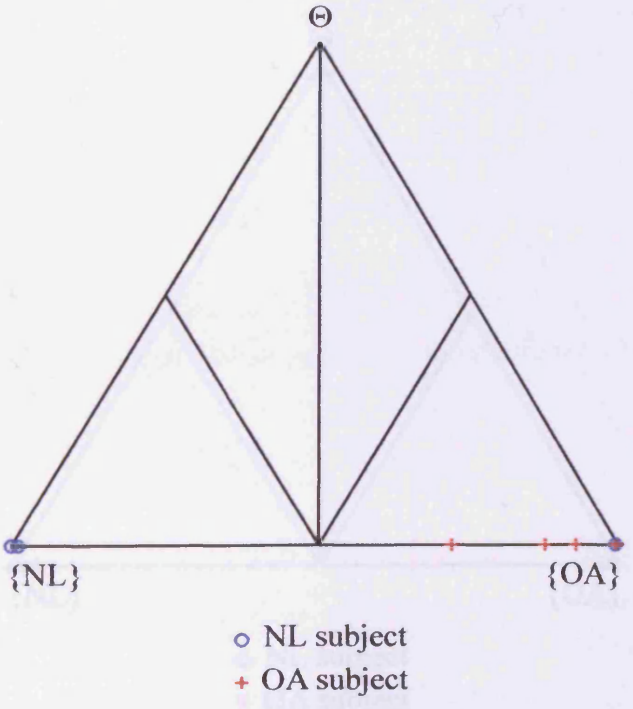


Figure B.41 Simplex plot showing simplex coordinates of out-of-sample subjects from test 4Ia ($[\Theta_L, \Theta_U] = [0, 0.7]$ and $k = k_c$).

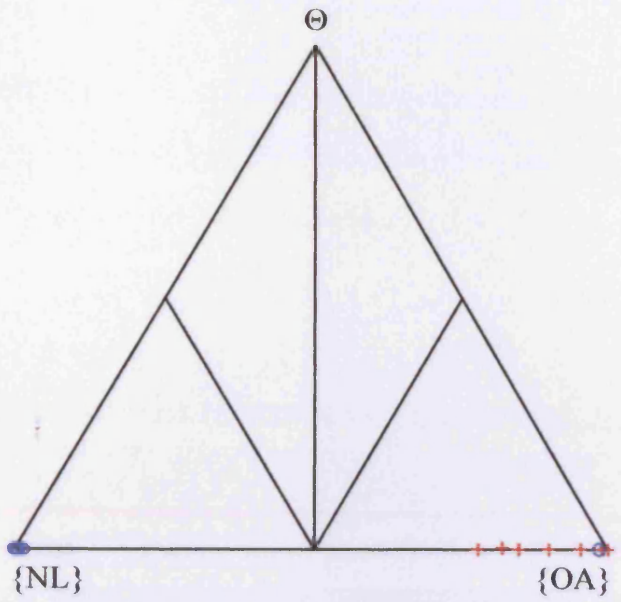


Figure B.42 Simplex plot showing simplex coordinates of out-of-sample subjects from test 4Ib ($[\Theta_L, \Theta_U] = [0.1, 0.7]$ and $k = k_c$).

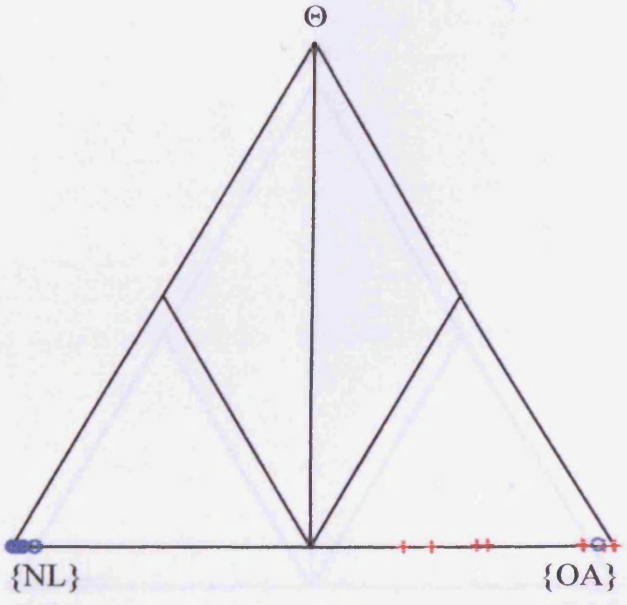


Figure B.43 Simplex plot showing simplex coordinates of out-of-sample subjects from test 4Ic ($[\Theta_L, \Theta_U] = [0.2, 0.7]$ and $k = k_c$).

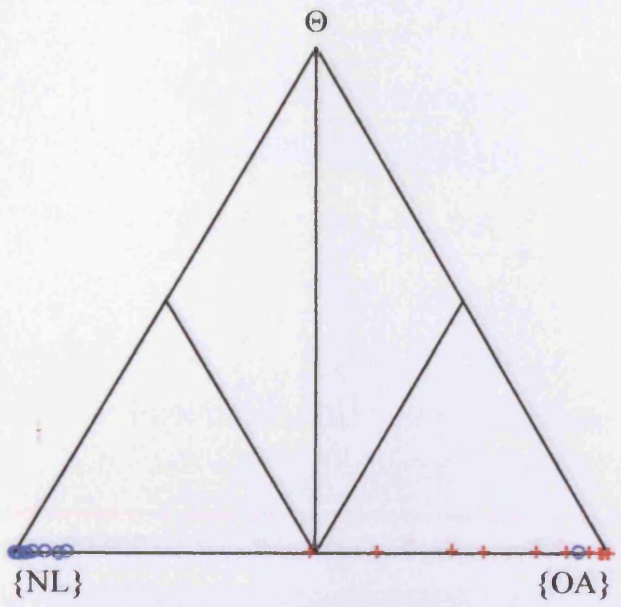


Figure B.44 Simplex plot showing simplex coordinates of out-of-sample subjects from test 4Id ($[\Theta_L, \Theta_U] = [0.3, 0.7]$ and $k = k_c$)

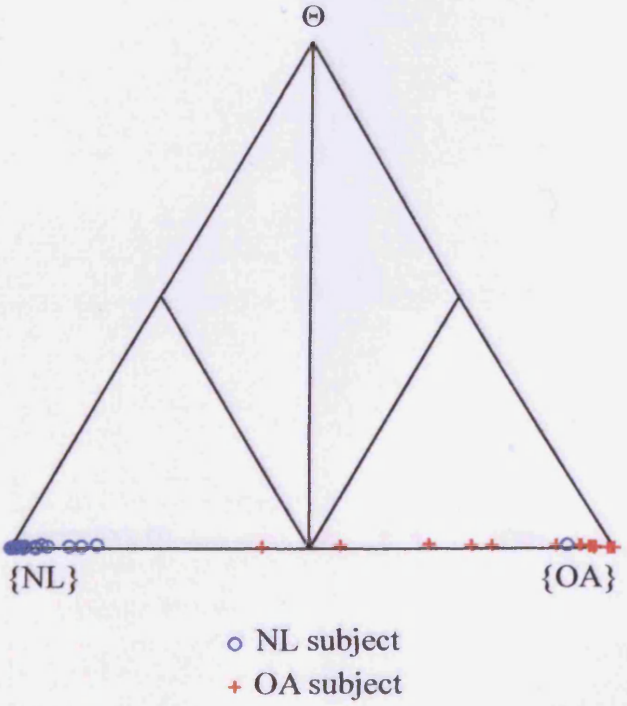


Figure B.45 Simplex plot showing simplex coordinates of out-of-sample subjects from test 4Ie ($[\Theta_L, \Theta_U] = [0.4, 0.7]$ and $k = k_c$)

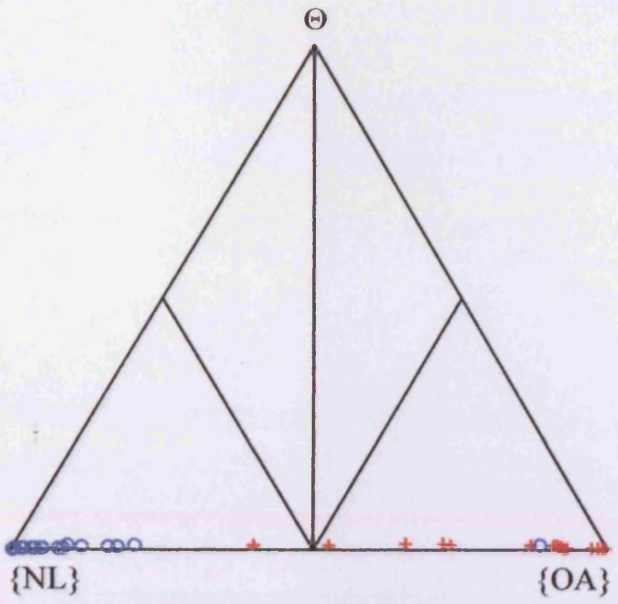


Figure B.46 Simplex plot showing simplex coordinates of out-of-sample subjects from test 4If ($[\Theta_L, \Theta_U] = [0.5, 0.7]$ and $k = k_c$)

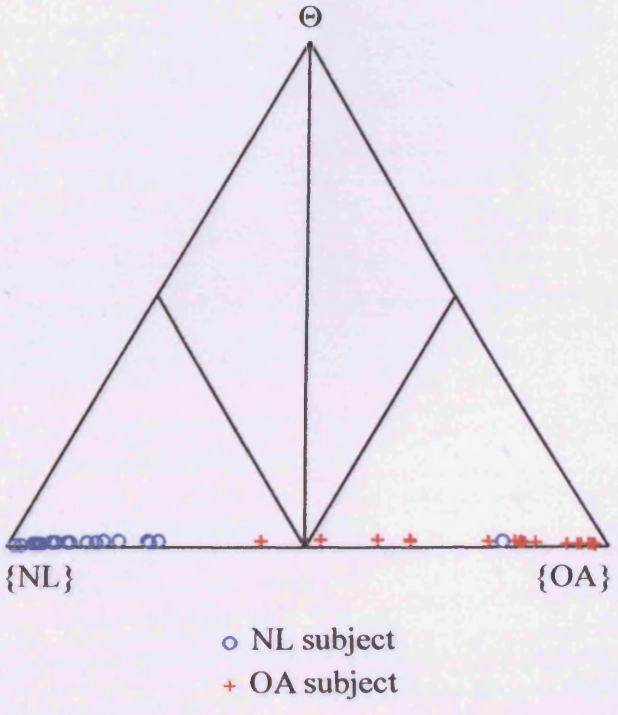


Figure B.47 Simplex plot showing simplex coordinates of out-of-sample subjects from test 4Ig ($[\Theta_L, \Theta_U] = [0.6, 0.7]$ and $k = k_c$)

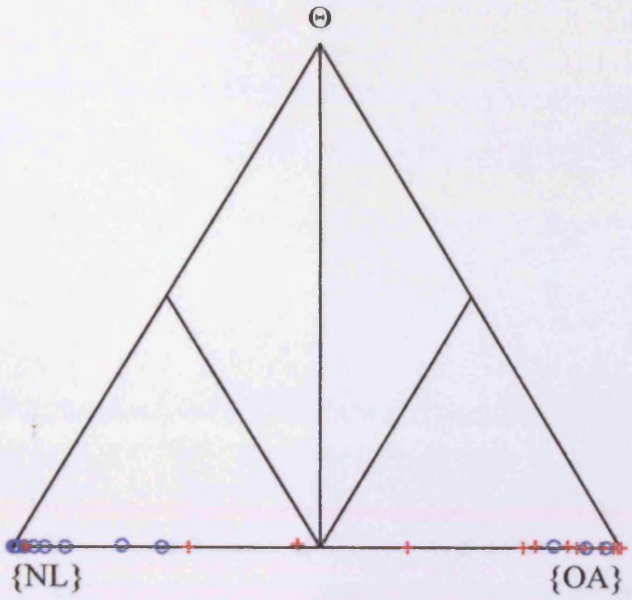


Figure B.48 Simplex plot showing simplex coordinates of out-of-sample subjects from test 4IIa ($[\Theta_L, \Theta_U] = [0, 0.7]$ and $k = k_s$)

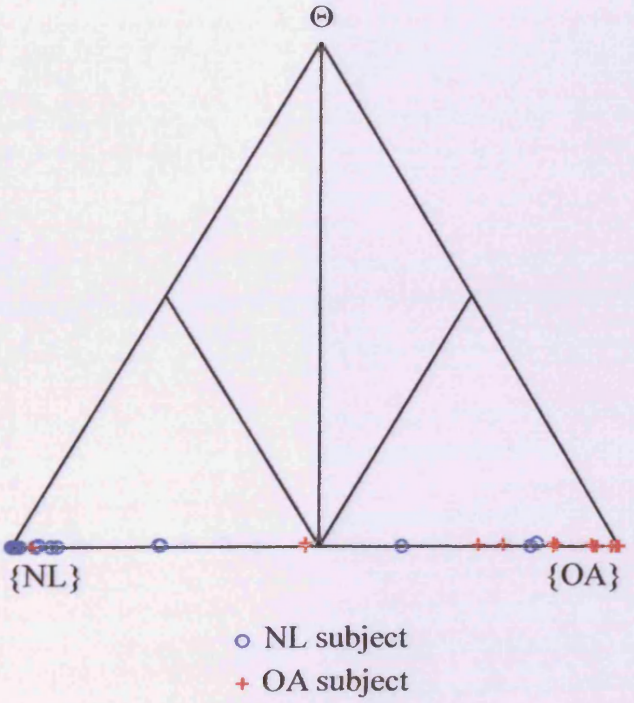


Figure B.49 Simplex plot showing simplex coordinates of out-of-sample subjects from test 4IIb ($[\Theta_L, \Theta_U] = [0.1, 0.7]$ and $k = k_s$)

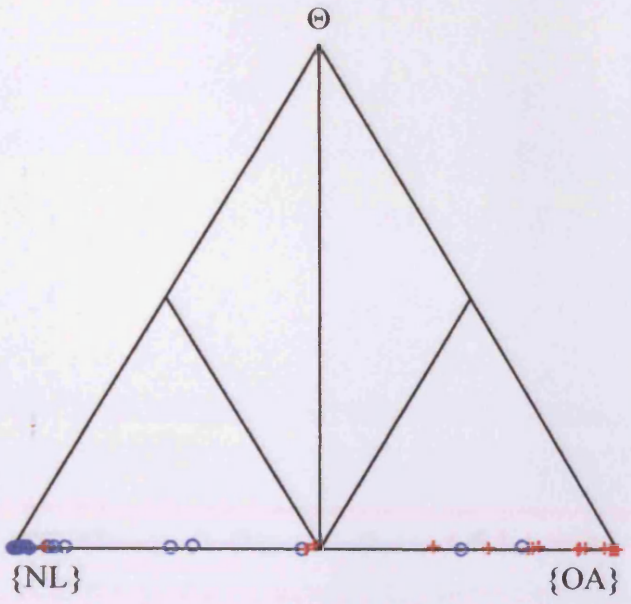


Figure B.50 Simplex plot showing simplex coordinates of out-of-sample subjects from test 4IIc ($[\Theta_L, \Theta_U] = [0.2, 0.7]$ and $k = k_s$)

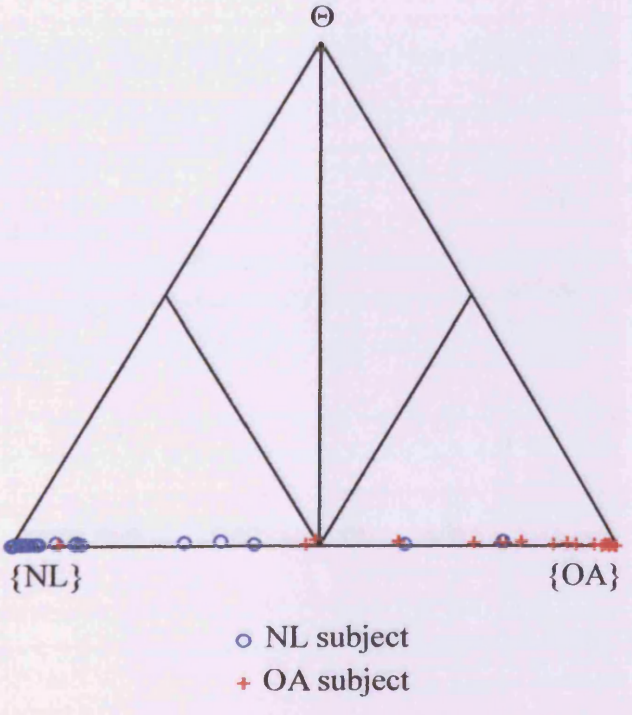


Figure B.51 Simplex plot showing simplex coordinates of out-of-sample subjects from test 4IId ($[\Theta_L, \Theta_U] = [0.3, 0.7]$ and $k = k_s$)

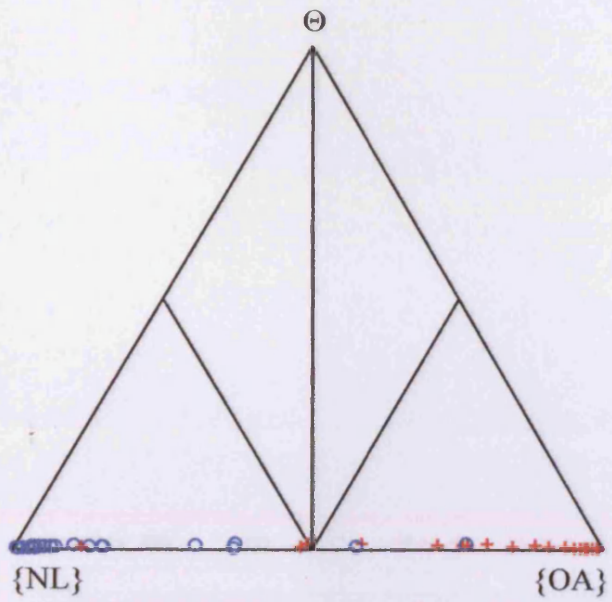


Figure B.52 Simplex plot showing simplex coordinates of out-of-sample subjects from test 4IIe ($[\Theta_L, \Theta_U] = [0.4, 0.7]$ and $k = k_s$)

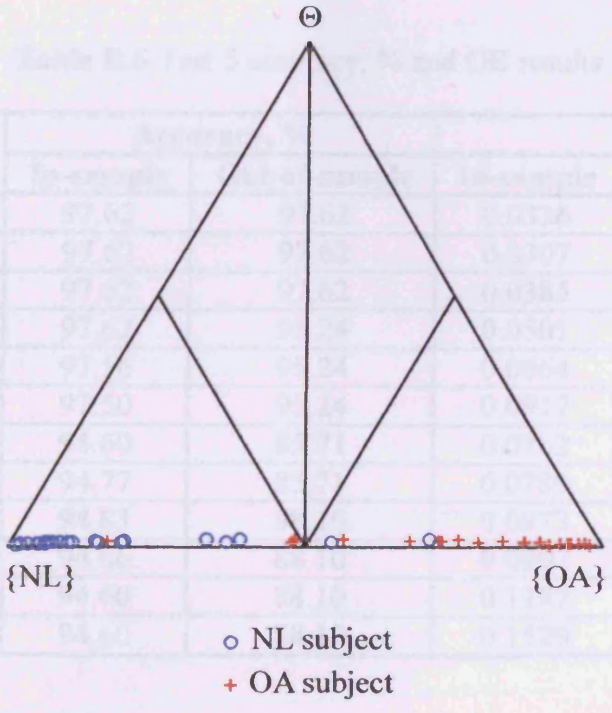


Figure B.53 Simplex plot showing simplex coordinates of out-of-sample subjects from test 4IIf ($[\Theta_L, \Theta_U] = [0.5, 0.7]$ and $k = k_s$)

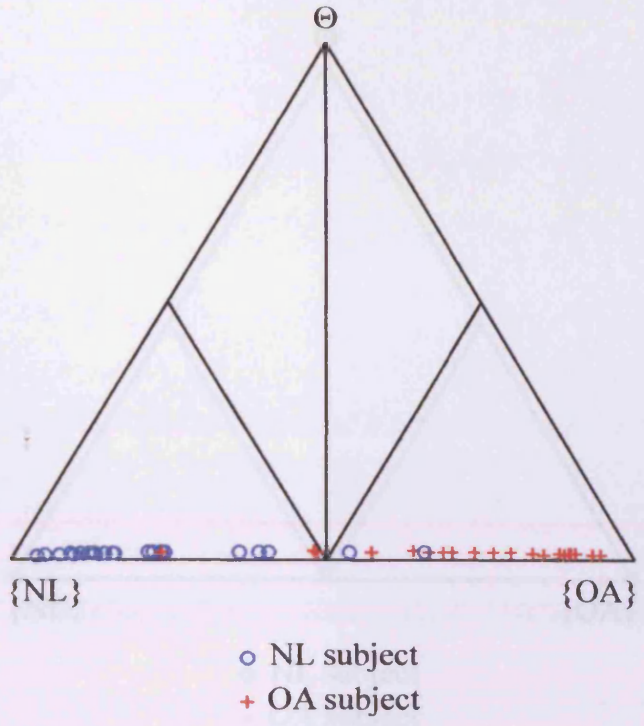


Figure B.54 Simplex plot showing simplex coordinates of out-of-sample subjects from test 4IIg ($[\Theta_L, \Theta_U] = [0.6, 0.7]$ and $k = k_s$)

Table B.6 Test 5 accuracy, % and OB results

Subtest	Part	Accuracy, %		OB	
		In-sample	Out-of-sample	In-sample	Out-of-sample
I	a	97.62	97.62	0.0336	0.0332
	b	97.62	97.62	0.0307	0.0415
	c	97.62	97.62	0.0385	0.0570
	d	97.62	95.24	0.0501	0.0731
	e	97.56	95.24	0.0664	0.0920
	f	97.50	95.24	0.0917	0.1190
II	a	94.60	85.71	0.0762	0.1545
	b	94.77	85.71	0.0789	0.1397
	c	94.83	88.10	0.0873	0.1441
	d	94.66	88.10	0.0997	0.1544
	e	94.60	88.10	0.1182	0.1711
	f	94.60	88.10	0.1529	0.2040

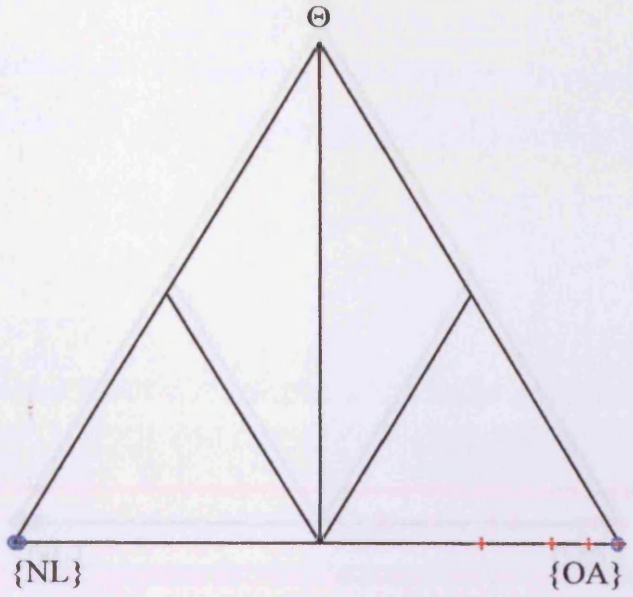


Figure B.55 Simplex plot showing simplex coordinates of out-of-sample subjects from test 5Ia ($[\Theta_L, \Theta_U] = [0, 0.6]$ and $k = k_c$).

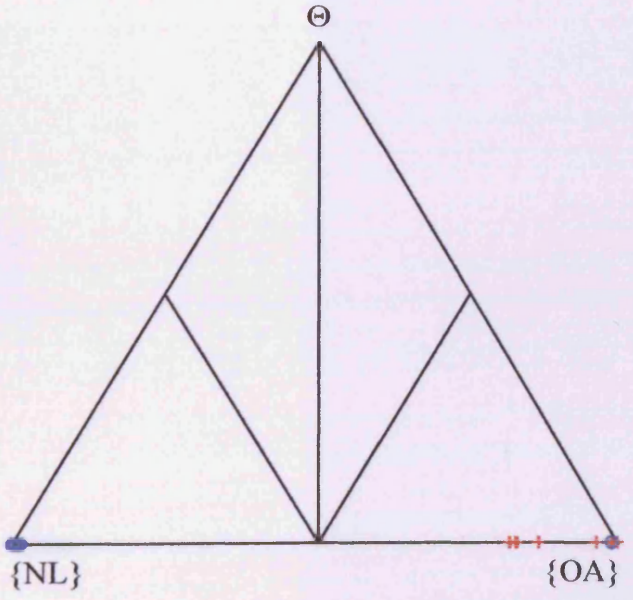


Figure B.56 Simplex plot showing simplex coordinates of out-of-sample subjects from test 5Ib ($[\Theta_L, \Theta_U] = [0.1, 0.6]$ and $k = k_c$).

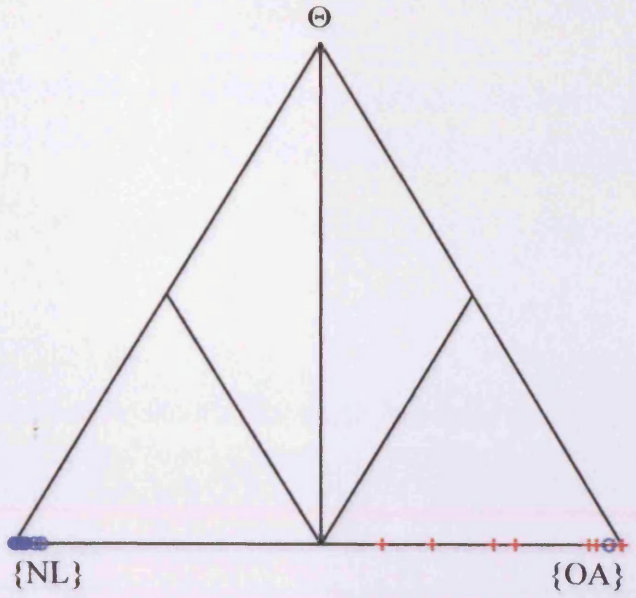


Figure B.57 Simplex plot showing simplex coordinates of out-of-sample subjects from test 5Ic ($[\Theta_L, \Theta_U] = [0.2, 0.6]$ and $k = k_c$).

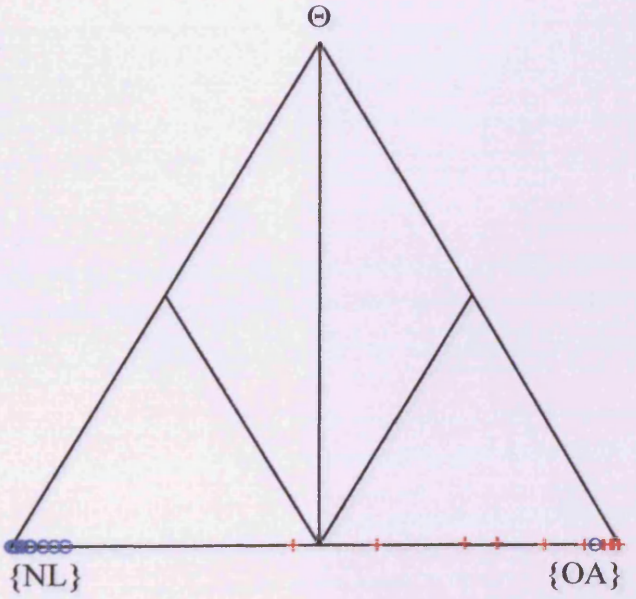


Figure B.58 Simplex plot showing simplex coordinates of out-of-sample subjects from test 5Id ($[\Theta_L, \Theta_U] = [0.3, 0.6]$ and $k = k_c$)

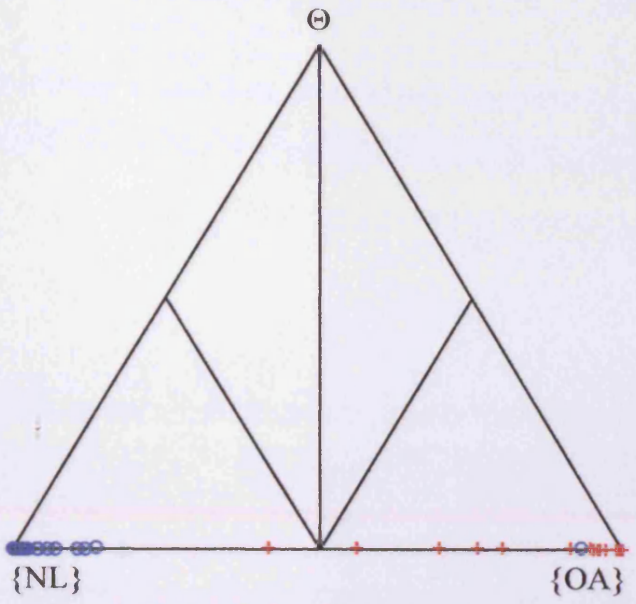


Figure B.59 Simplex plot showing simplex coordinates of out-of-sample subjects from test 5Ie ($[\Theta_L, \Theta_U] = [0.4, 0.6]$ and $k = k_c$)

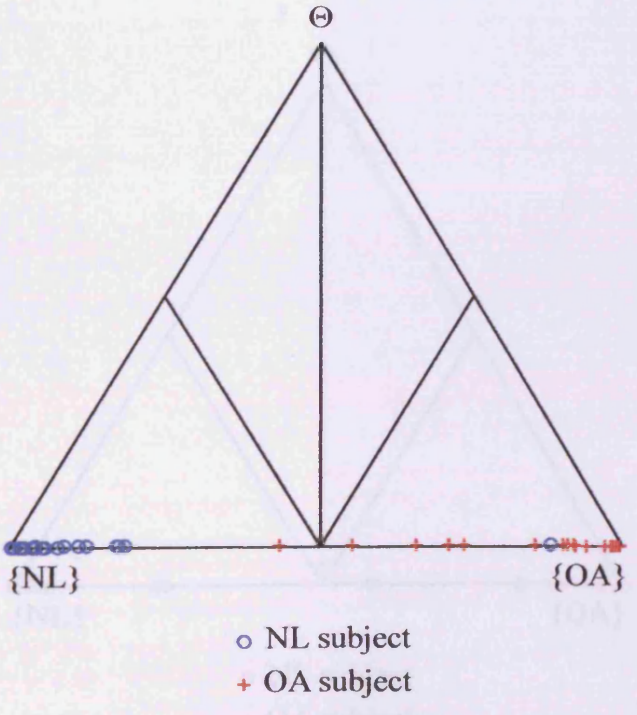


Figure B.60 Simplex plot showing simplex coordinates of out-of-sample subjects from test 5If ($[\Theta_L, \Theta_U] = [0.5, 0.6]$ and $k = k_c$).

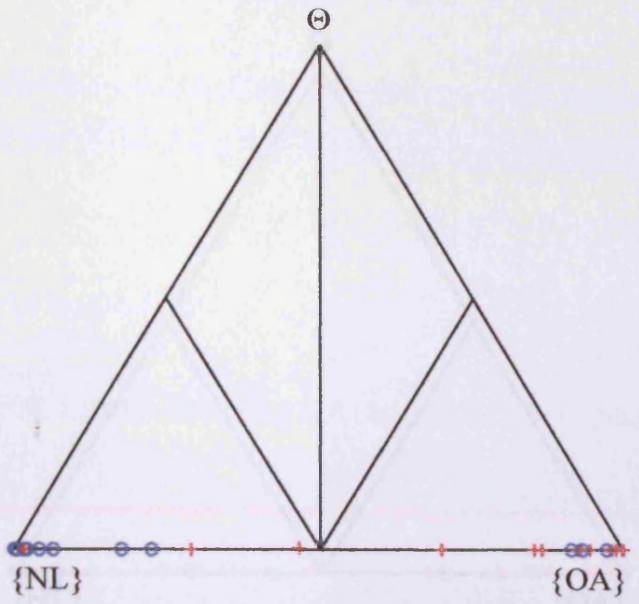


Figure B.61 Simplex plot showing simplex coordinates of out-of-sample subjects from test 5IIa ($[\Theta_L, \Theta_U] = [0, 0.6]$ and $k = k_s$)

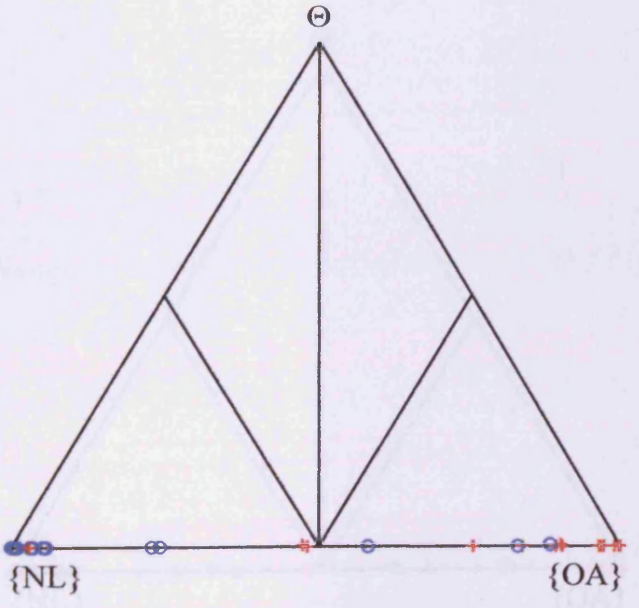


Figure B.62 Simplex plot showing simplex coordinates of out-of-sample subjects from test 5IIb ($[\Theta_L, \Theta_U] = [0.1, 0.6]$ and $k = k_s$)

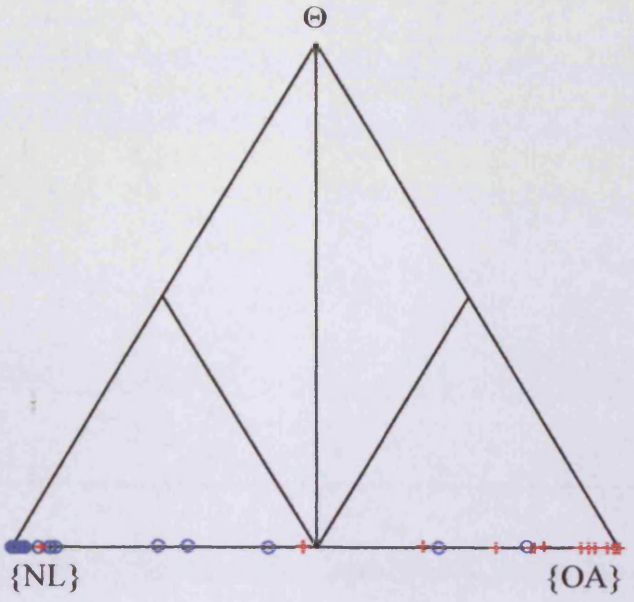


Figure B.63 Simplex plot showing simplex coordinates of out-of-sample subjects from test 5IIc ($[\Theta_L, \Theta_U] = [0.2, 0.6]$ and $k = k_s$)

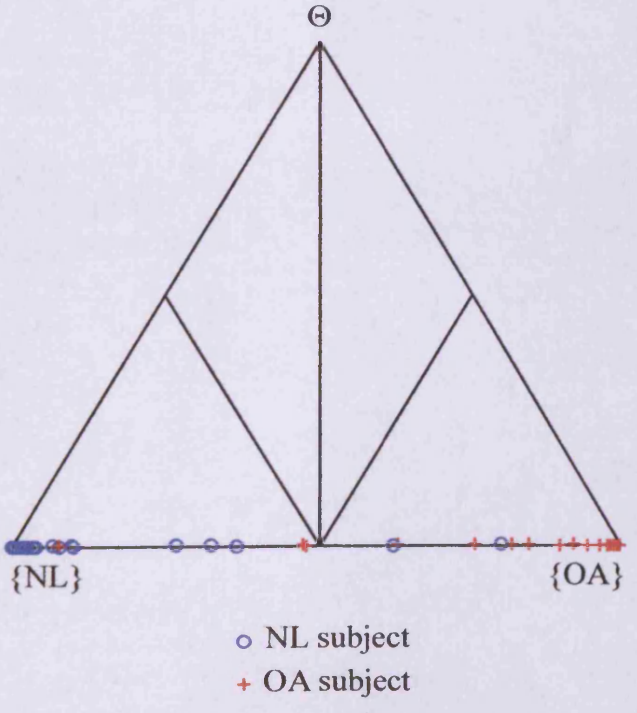


Figure B.64 Simplex plot showing simplex coordinates of out-of-sample subjects from test 5IId ($[\Theta_L, \Theta_U] = [0.3, 0.6]$ and $k = k_s$)

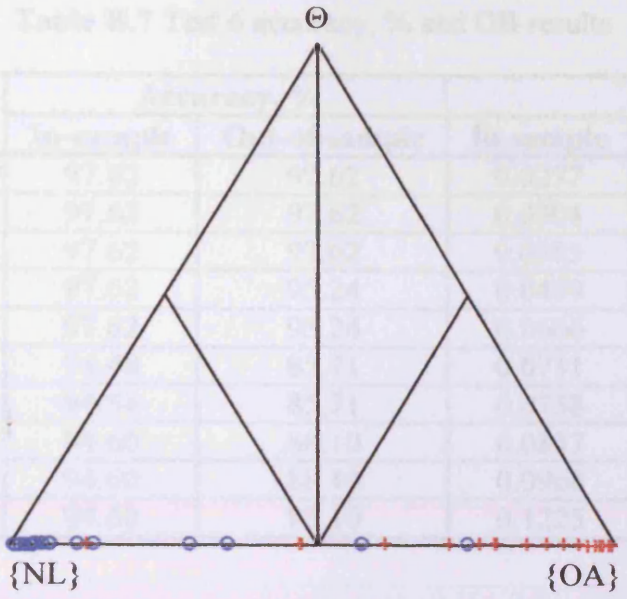


Figure B.65 Simplex plot showing simplex coordinates of out-of-sample subjects from test 5IIe ($[\Theta_L, \Theta_U] = [0.4, 0.6]$ and $k = k_s$)

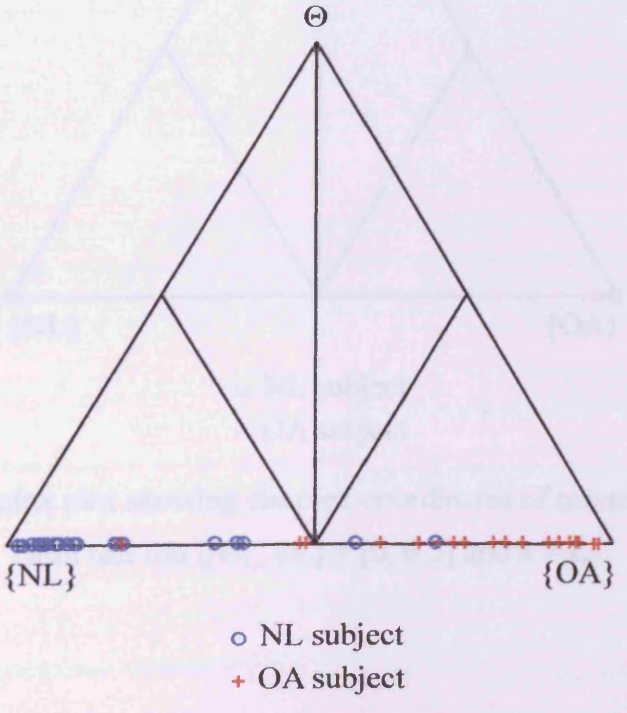


Figure B.66 Simplex plot showing simplex coordinates of out-of-sample subjects from test 5IIf ($[\Theta_L, \Theta_U] = [0.5, 0.6]$ and $k = k_s$)

Table B.7 Test 6 accuracy, % and OB results

Subtest	Part	Accuracy, %		OB	
		In-sample	Out-of-sample	In-sample	Out-of-sample
I	a	97.62	97.62	0.0277	0.0318
	b	97.62	97.62	0.0304	0.0417
	c	97.62	97.62	0.0385	0.0577
	d	97.62	95.24	0.0499	0.0725
	e	97.62	95.24	0.0666	0.0914
II	a	94.48	85.71	0.0731	0.1501
	b	94.54	85.71	0.0758	0.1358
	c	94.60	88.10	0.0837	0.1400
	d	94.60	88.10	0.0968	0.1512
	e	94.60	88.10	0.1225	0.1760

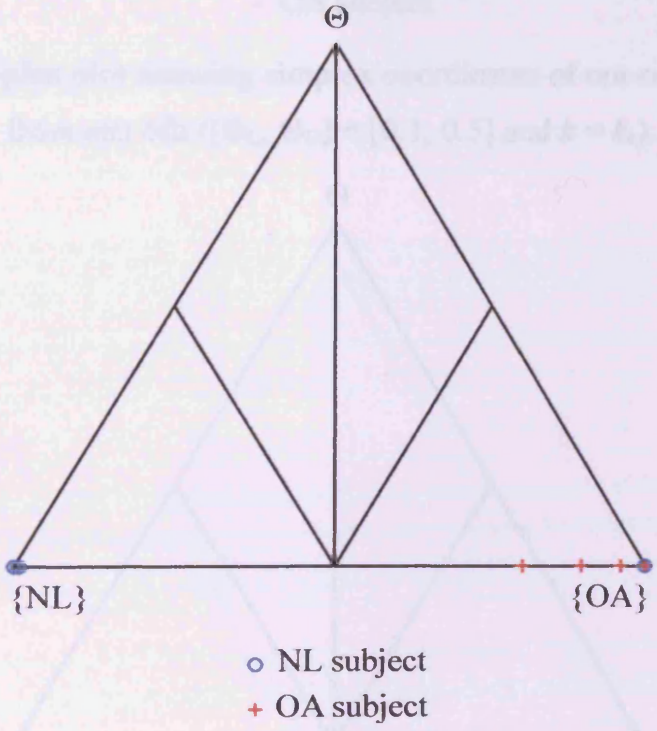


Figure B.67 Simplex plot showing simplex coordinates of out-of-sample subjects from test 6Ia ($[\Theta_L, \Theta_U] = [0, 0.5]$ and $k = k_c$).

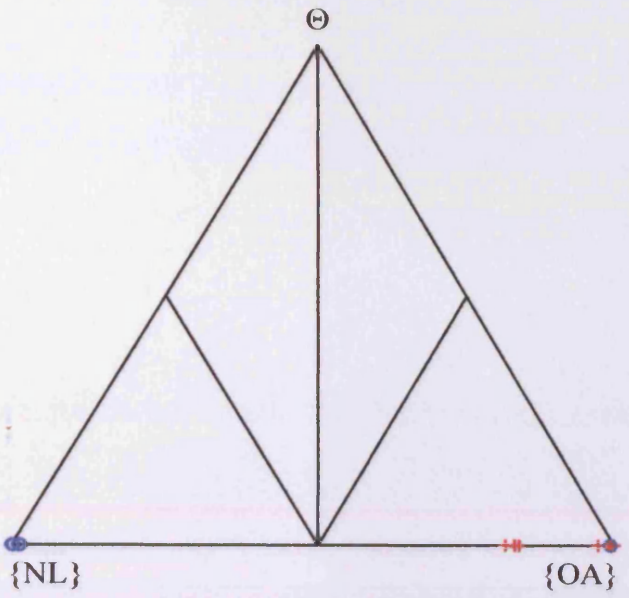


Figure B.68 Simplex plot showing simplex coordinates of out-of-sample subjects from test 6Ib ($[\Theta_L, \Theta_U] = [0.1, 0.5]$ and $k = k_c$).

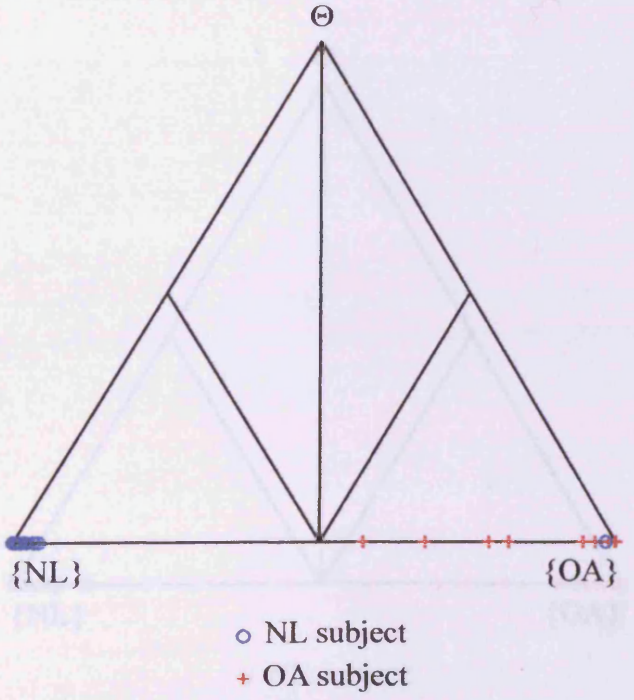


Figure B.69 Simplex plot showing simplex coordinates of out-of-sample subjects from test 6Ic ($[\Theta_L, \Theta_U] = [0.2, 0.5]$ and $k = k_c$).

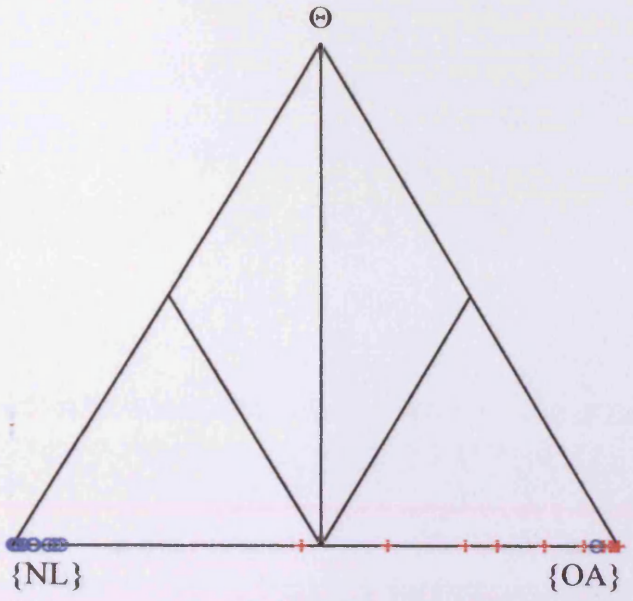


Figure B.70 Simplex plot showing simplex coordinates of out-of-sample subjects from test 6Id ($[\Theta_L, \Theta_U] = [0.3, 0.5]$ and $k = k_c$)

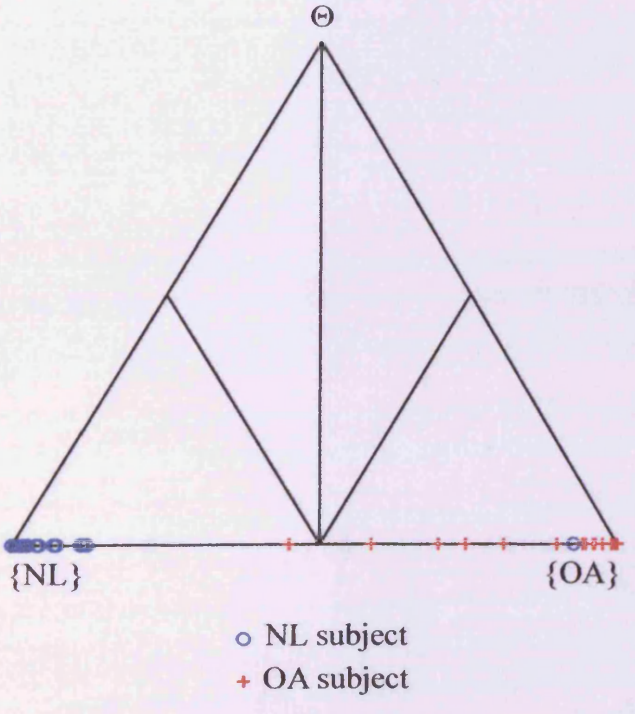


Figure B.71 Simplex plot showing simplex coordinates of out-of-sample subjects from test 6Ie ($[\Theta_L, \Theta_U] = [0.4, 0.5]$ and $k = k_c$)

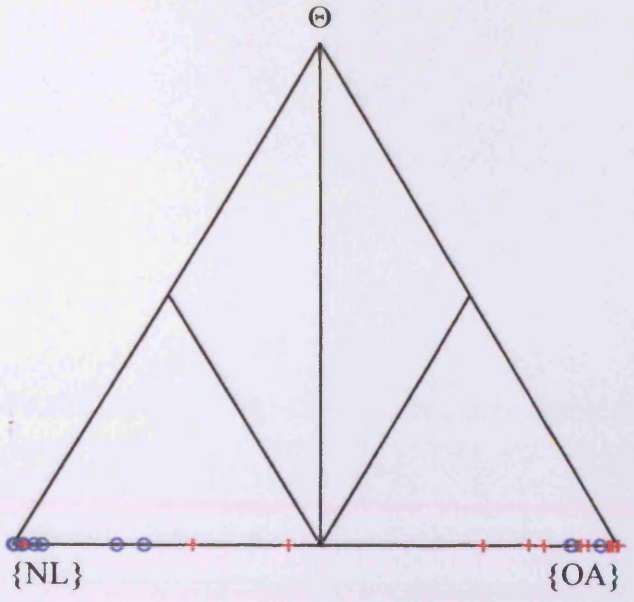


Figure B.72 Simplex plot showing simplex coordinates of out-of-sample subjects from test 6IIa ($[\Theta_L, \Theta_U] = [0, 0.5]$ and $k = k_s$)

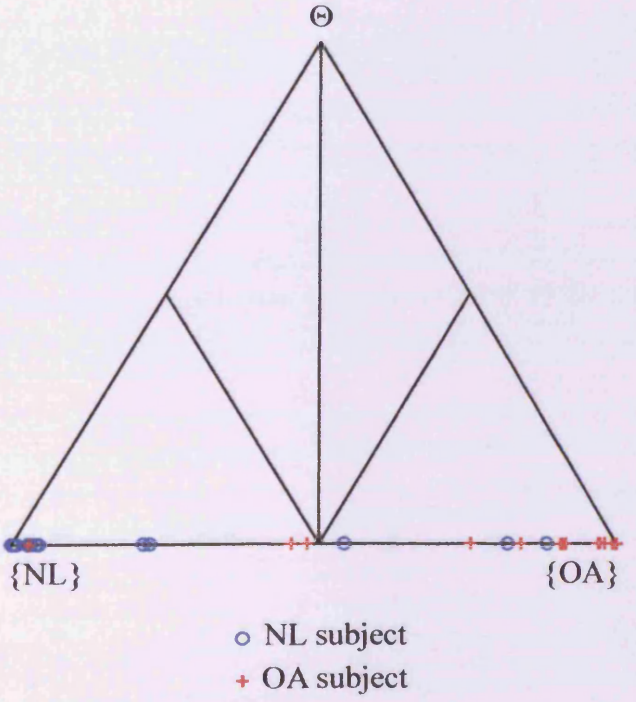


Figure B.73 Simplex plot showing simplex coordinates of out-of-sample subjects from test 6IIb ($[\Theta_L, \Theta_U] = [0.1, 0.5]$ and $k = k_s$)

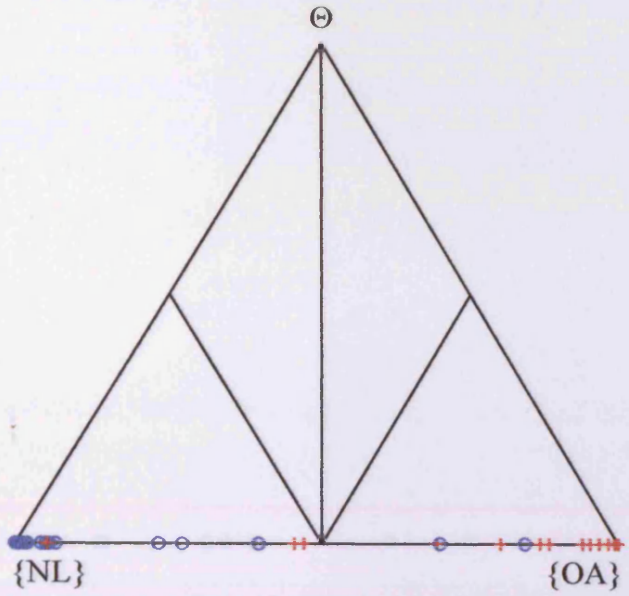


Figure B.74 Simplex plot showing simplex coordinates of out-of-sample subjects from test 6IIc ($[\Theta_L, \Theta_U] = [0.2, 0.5]$ and $k = k_s$)

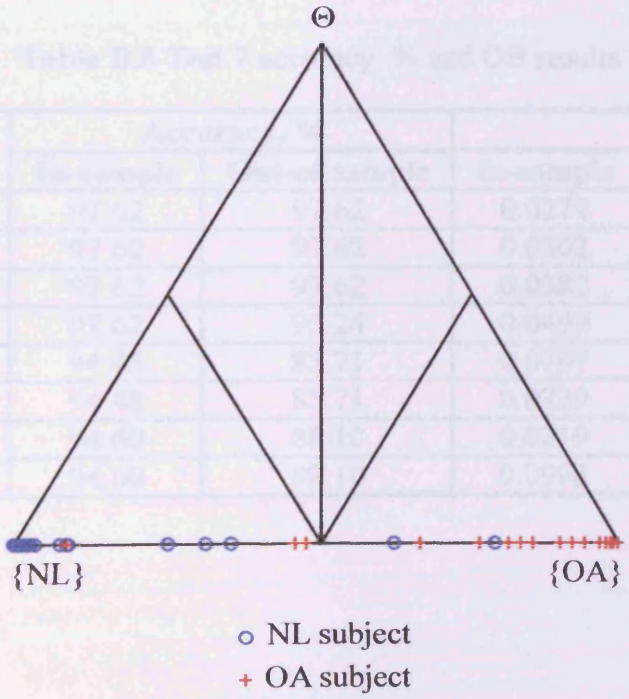


Figure B.75 Simplex plot showing simplex coordinates of out-of-sample subjects from test 6IId ($[\Theta_L, \Theta_U] = [0.3, 0.5]$ and $k = k_s$)

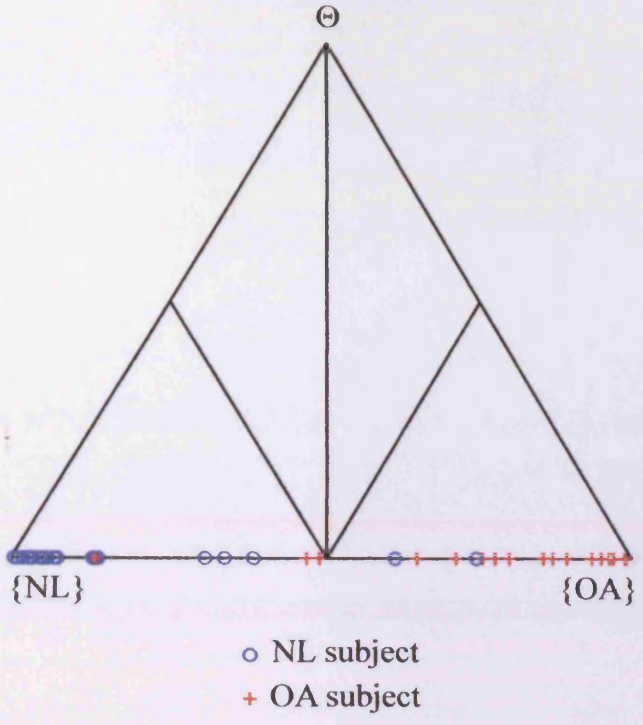


Figure B.76 Simplex plot showing simplex coordinates of out-of-sample subjects from test 6IIe ($[\Theta_L, \Theta_U] = [0.4, 0.5]$ and $k = k_s$)

Table B.8 Test 7 accuracy, % and OB results

Subtest	Part	Accuracy, %		OB	
		In-sample	Out-of-sample	In-sample	Out-of-sample
I	a	97.62	97.62	0.0274	0.0314
	b	97.62	97.62	0.0302	0.0417
	c	97.62	97.62	0.0382	0.0568
	d	97.62	95.24	0.0499	0.0717
II	a	94.43	85.71	0.0707	0.1469
	b	94.48	85.71	0.0729	0.1324
	c	94.60	88.10	0.0819	0.1380
	d	94.60	88.10	0.0998	0.1552

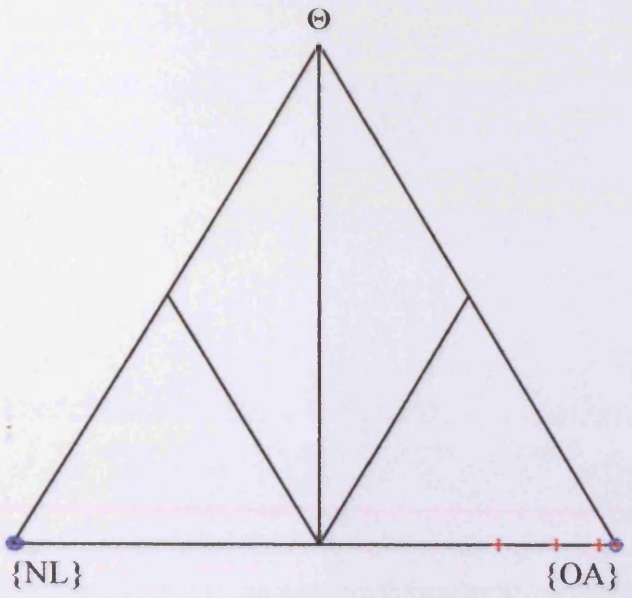


Figure B.77 Simplex plot showing simplex coordinates of out-of-sample subjects from test 7Ia ($[\Theta_L, \Theta_U] = [0, 0.4]$ and $k = k_c$).

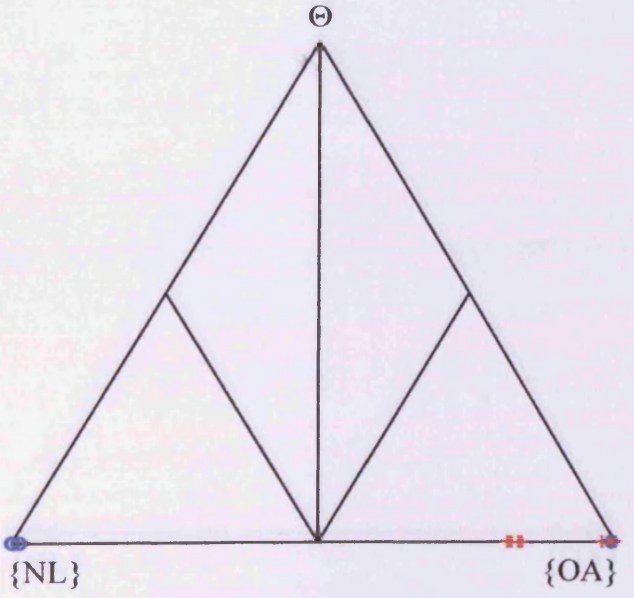


Figure 4.78 Simplex plot showing simplex coordinates of out-of-sample subjects from test 7Ib ($[\Theta_L, \Theta_U] = [0.1, 0.4]$ and $k = k_c$).

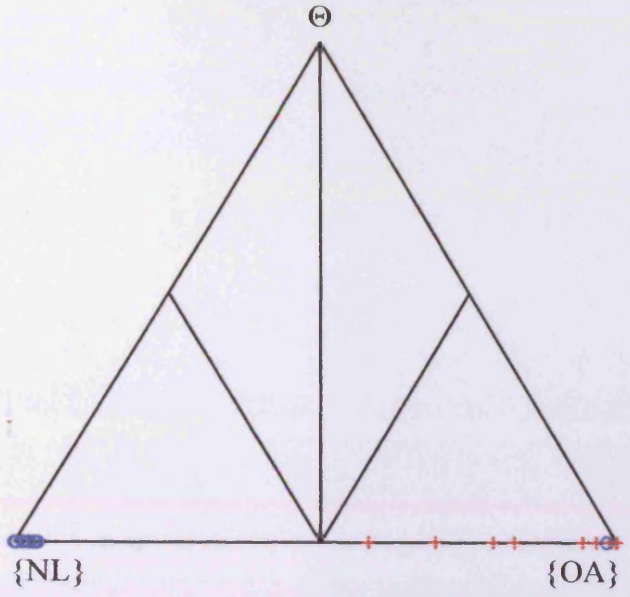


Figure 4.79 Simplex plot showing simplex coordinates of out-of-sample subjects from test 7Ic ($[\Theta_L, \Theta_U] = [0.2, 0.4]$ and $k = k_c$).

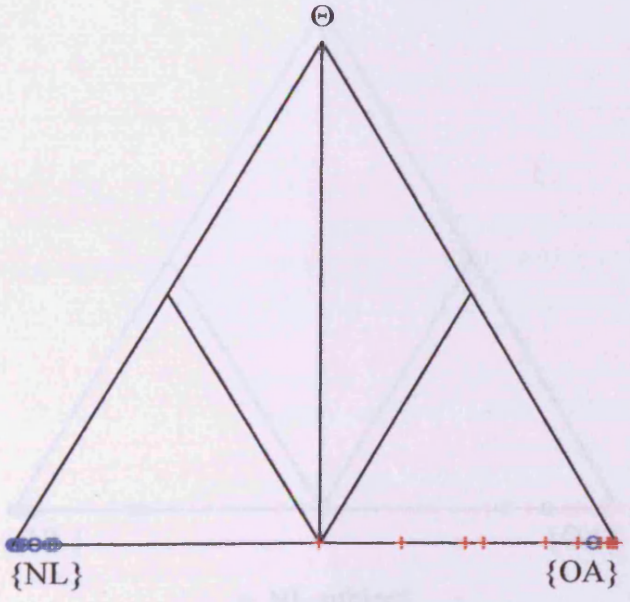
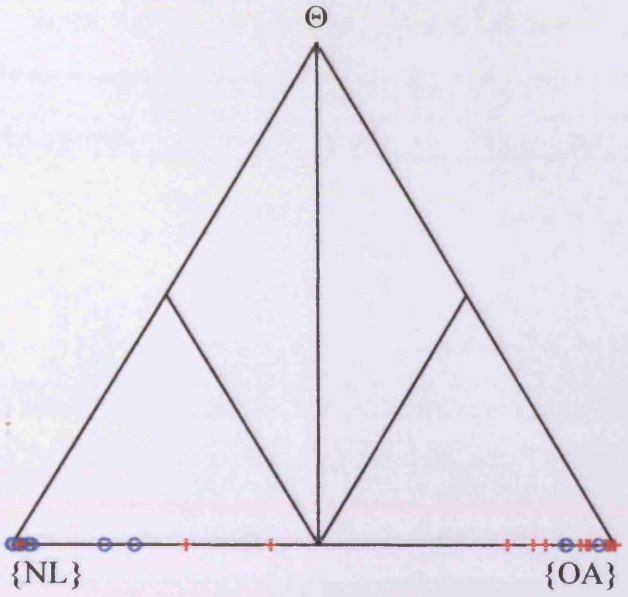
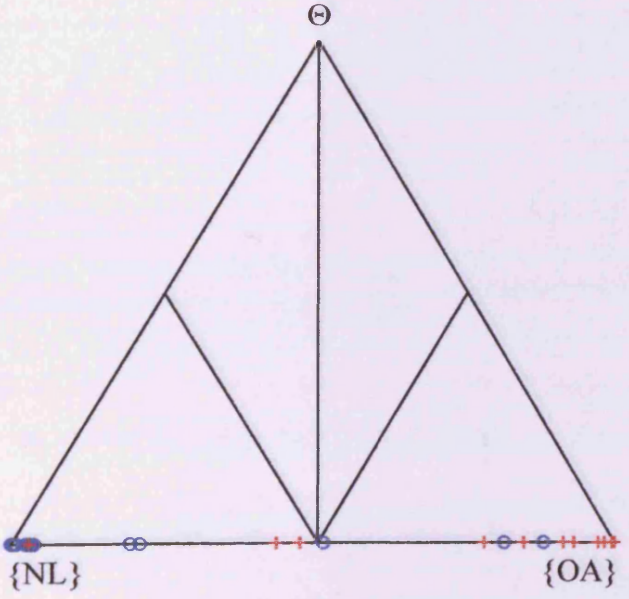


Figure B.80 Simplex plot showing simplex coordinates of out-of-sample subjects from test 7Id ($[\Theta_L, \Theta_U] = [0.3, 0.4]$ and $k = k_c$)



o NL subject
+ OA subject

Figure B.81 Simplex plot showing simplex coordinates of out-of-sample subjects from test 7IIa ($[\Theta_L, \Theta_U] = [0, 0.4]$ and $k = k_s$)



o NL subject
+ OA subject

Figure B.82 Simplex plot showing simplex coordinates of out-of-sample subjects from test 7IIb ($[\Theta_L, \Theta_U] = [0.1, 0.4]$ and $k = k_s$)

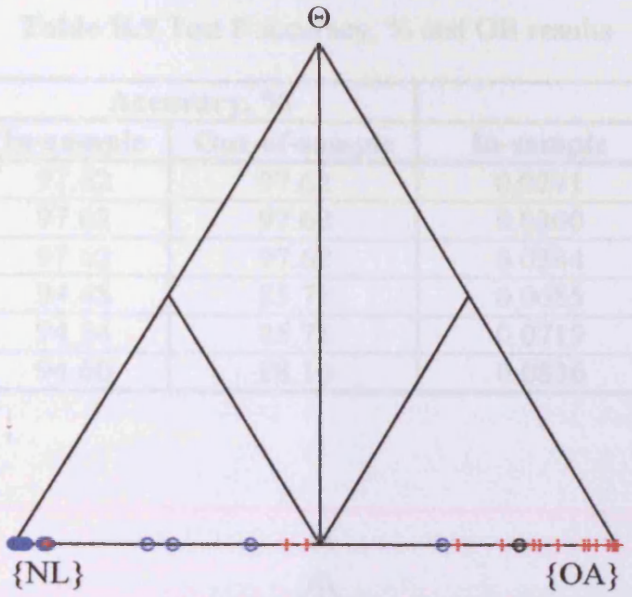


Figure B.83 Simplex plot showing simplex coordinates of out-of-sample subjects from test 7IIc ($[\Theta_L, \Theta_U] = [0.2, 0.4]$ and $k = k_s$)

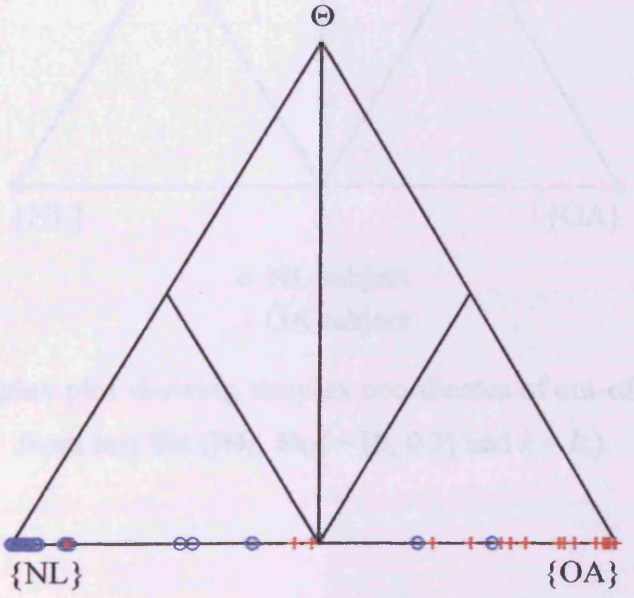


Figure B.84 Simplex plot showing simplex coordinates of out-of-sample subjects from test 7IId ($[\Theta_L, \Theta_U] = [0.3, 0.4]$ and $k = k_s$)

Table B.9 Test 8 accuracy, % and OB results

Subtest	Part	Accuracy, %		OB	
		In-sample	Out-of-sample	In-sample	Out-of-sample
I	a	97.62	97.62	0.0271	0.0309
	b	97.62	97.62	0.0300	0.0412
	c	97.62	97.62	0.0384	0.0564
II	a	94.48	85.71	0.0685	0.1440
	b	94.54	85.71	0.0719	0.1313
	c	94.60	88.10	0.0836	0.1409

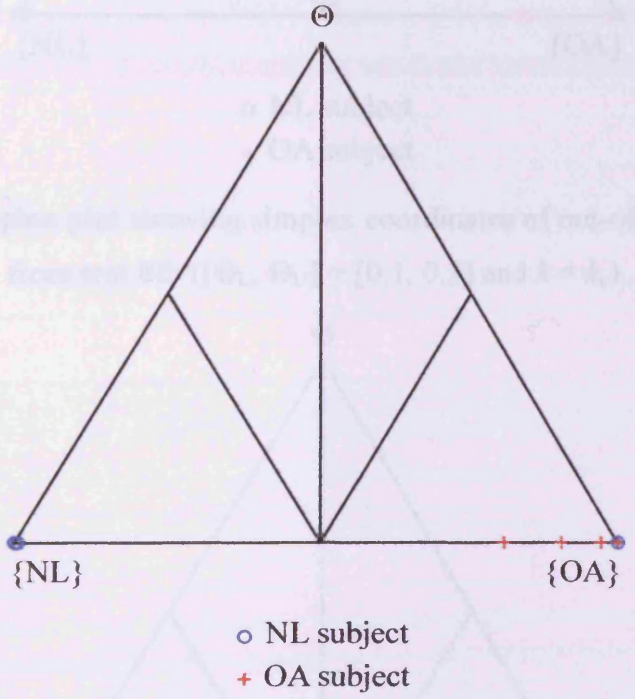


Figure 4.85 Simplex plot showing simplex coordinates of out-of-sample subjects from test 8Ia ($[\Theta_L, \Theta_U] = [0, 0.3]$ and $k = k_c$).

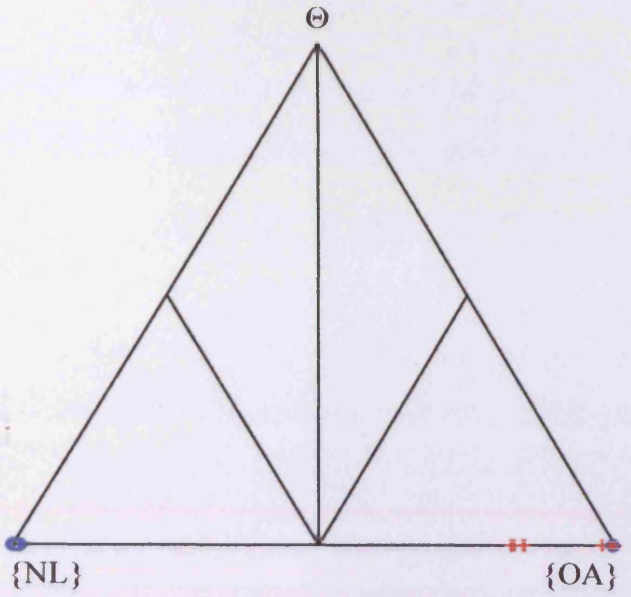


Figure 4.86 Simplex plot showing simplex coordinates of out-of-sample subjects from test 8Ib ($[\Theta_L, \Theta_U] = [0.1, 0.3]$ and $k = k_c$).

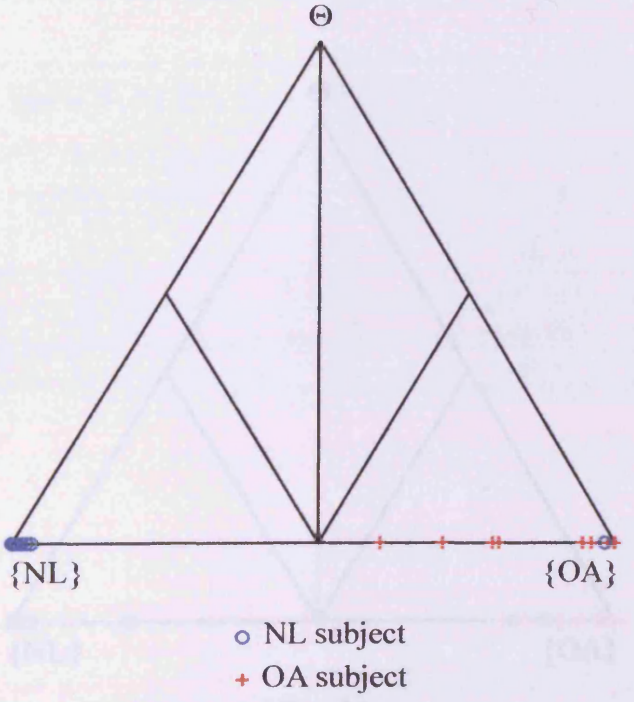


Figure 4.87 Simplex plot showing simplex coordinates of out-of-sample subjects from test 8Ic ($[\Theta_L, \Theta_U] = [0.2, 0.3]$ and $k = k_c$).

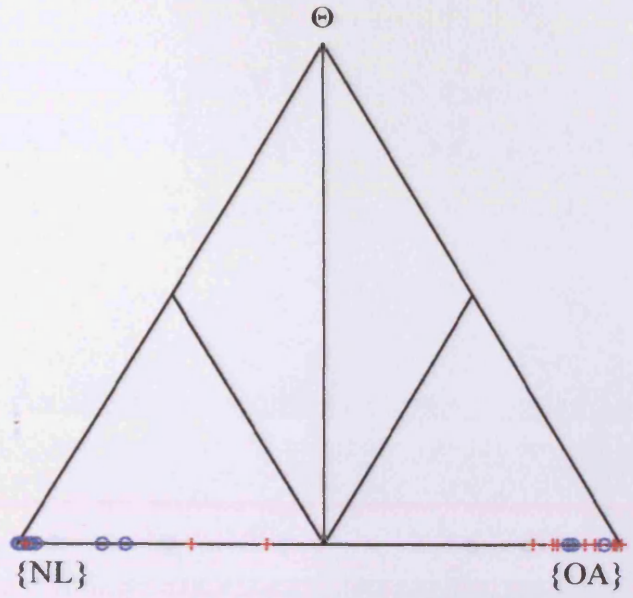


Figure B.88 Simplex plot showing simplex coordinates of out-of-sample subjects from test 8IIa ($[\Theta_L, \Theta_U] = [0, 0.3]$ and $k = k_s$)

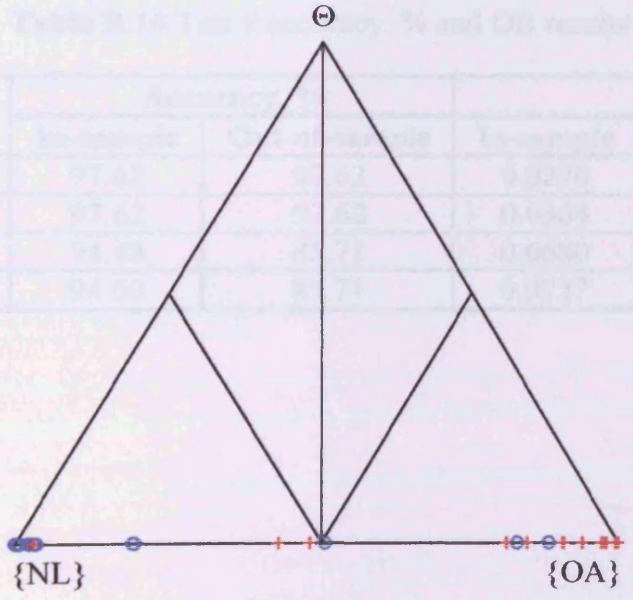


Figure B.89 Simplex plot showing simplex coordinates of out-of-sample subjects from test 8IIb ($[\Theta_L, \Theta_U] = [0.1, 0.3]$ and $k = k_s$)

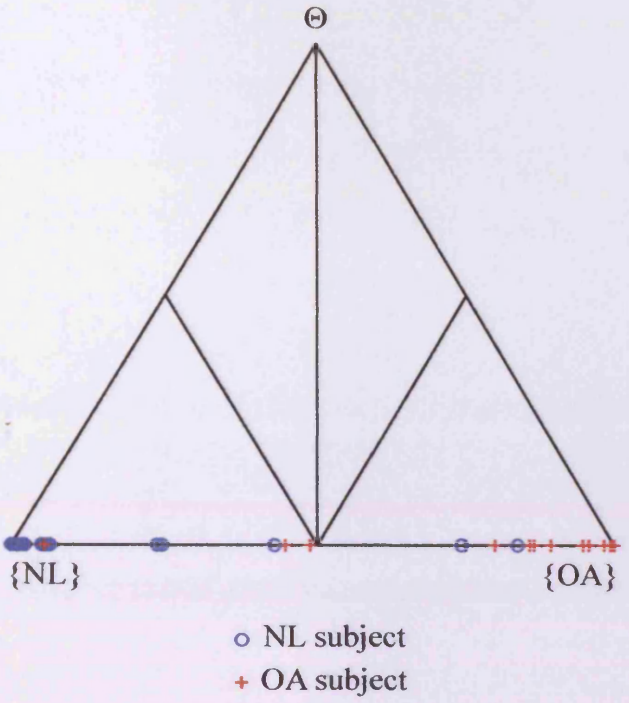


Figure B.90 Simplex plot showing simplex coordinates of out-of-sample subjects from test 8IIc ($[\Theta_L, \Theta_U] = [0.2, 0.3]$ and $k = k_s$)

Table B.10 Test 9 accuracy, % and OB results

Subtest	Part	Accuracy, %		OB	
		In-sample	Out-of-sample	In-sample	Out-of-sample
I	a	97.62	97.62	0.0270	0.0306
	b	97.62	97.62	0.0304	0.0414
II	a	94.43	85.71	0.0680	0.1428
	b	94.60	85.71	0.0727	0.1330

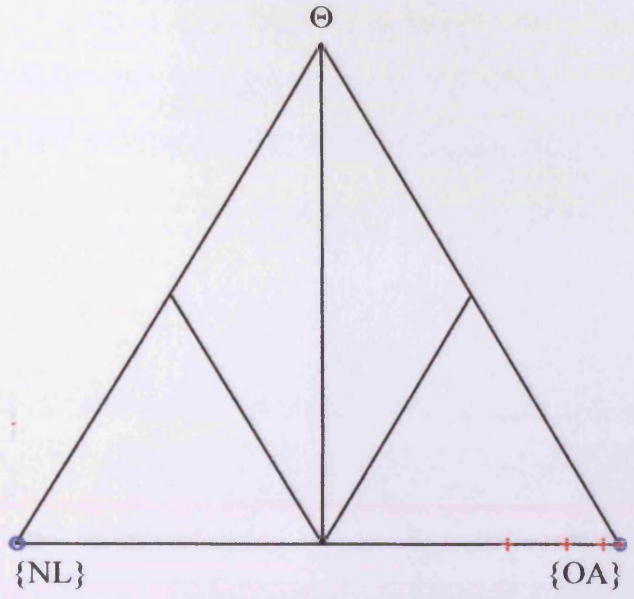


Figure B.91 Simplex plot showing simplex coordinates of out-of-sample subjects from test 9Ia ($[\Theta_L, \Theta_U] = [0, 0.2]$ and $k = k_c$).

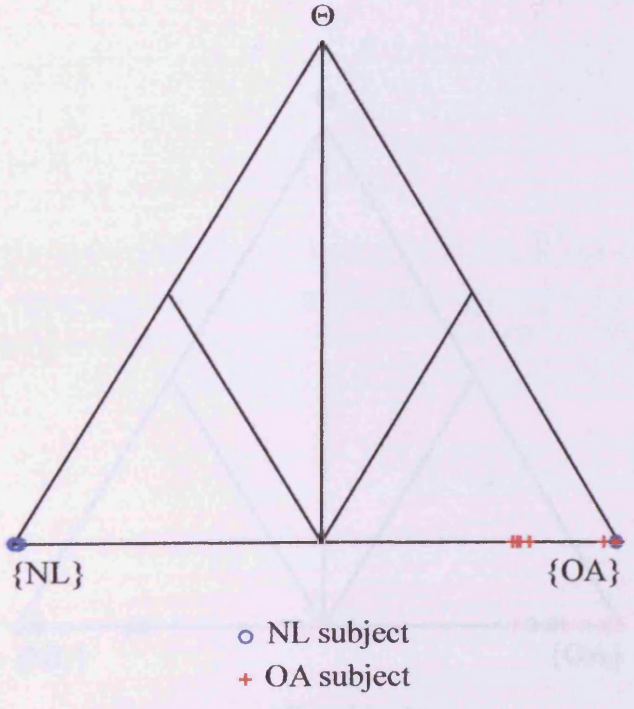


Figure B.92 Simplex plot showing simplex coordinates of out-of-sample subjects from test 9Ib ($[\Theta_L, \Theta_U] = [0.1, 0.2]$ and $k = k_c$).

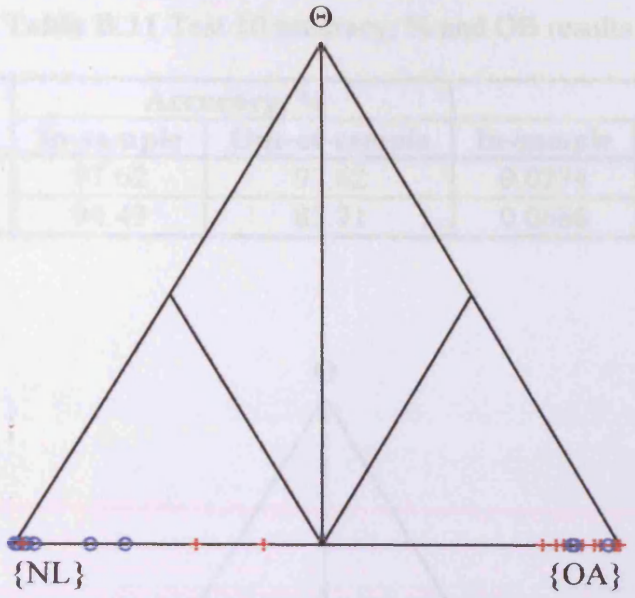


Figure B.93 Simplex plot showing simplex coordinates of out-of-sample subjects from test 9IIa ($[\Theta_L, \Theta_U] = [0, 0.2]$ and $k = k_s$)

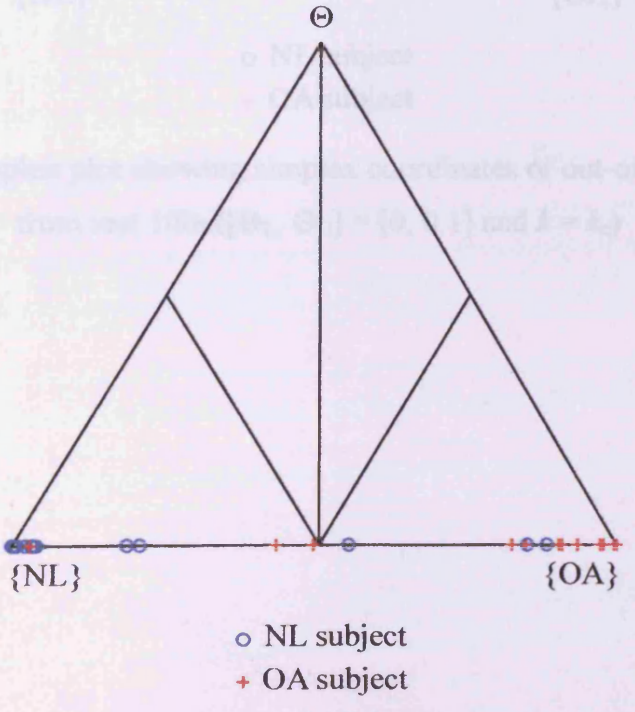


Figure B.94 Simplex plot showing simplex coordinates of out-of-sample subjects from test 9IIb ($[\Theta_L, \Theta_U] = [0.1, 0.2]$ and $k = k_s$)

Table B.11 Test 10 accuracy, % and OB results

Subtest	Part	Accuracy, %		OB	
		In-sample	Out-of-sample	In-sample	Out-of sample
I	a	97.62	97.62	0.0274	0.0312
II		94.43	85.71	0.0686	0.1426

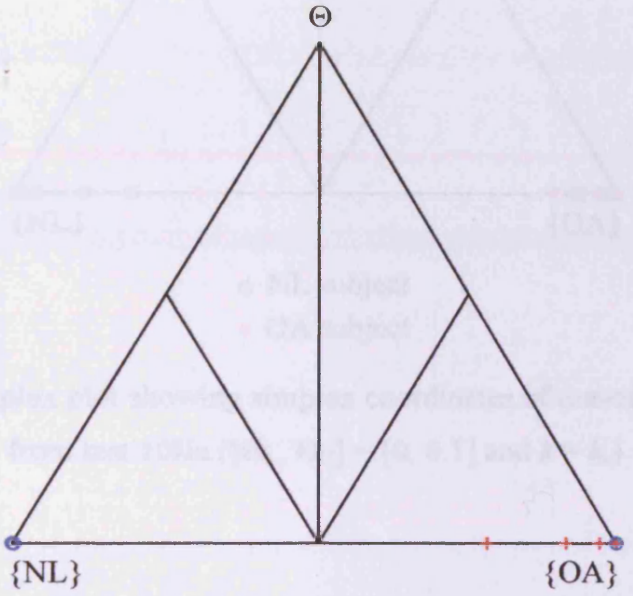


Figure B.95 Simplex plot showing simplex coordinates of out-of-sample subjects from test 10Ia ($[\Theta_L, \Theta_U] = [0, 0.1]$ and $k = k_c$)

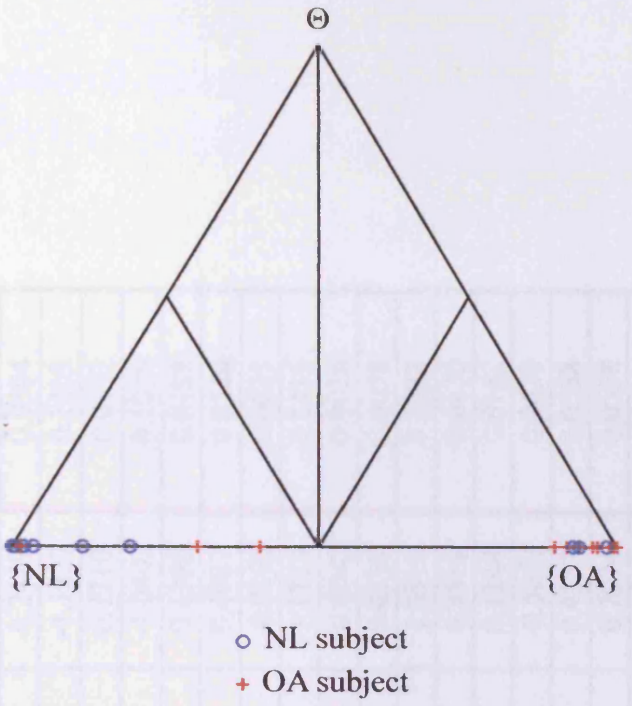


Figure B.96 Simplex plot showing simplex coordinates of out-of-sample subjects from test 10IIa ($[\Theta_L, \Theta_U] = [0, 0.1]$ and $k = k_s$)

Table B.12 Summary of average k values for the 10 runs of test 14.

Input Variable	Test					
	14a		14b		14c	
	$t_f = 1\times10^{-5}$		$t_f = 1\times10^{-10}$		$t_f = 1\times10^{-20}$	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
v_1	0.1761	0.0100	0.1891	0.0282	0.2449	0.0504
v_2	-0.7879	0.0511	-0.9018	0.0748	-1.1127	0.1403
v_3	0.6299	0.0425	0.6649	0.0930	0.7710	0.0557
v_4	-0.8318	0.0557	-0.8759	0.0800	-1.0403	0.1026
v_5	0.1076	0.0117	0.1280	0.0108	0.1396	0.0188
v_6	0.3676	0.0250	0.4128	0.0707	0.4907	0.0349
v_7	-0.0240	0.0018	-0.0235	0.0051	-0.0192	0.0025
v_8	0.9526	0.0760	0.9970	0.0830	1.1713	0.1415
v_9	-0.0045	0.0006	-0.0046	0.0007	-0.0045	0.0016
v_{10}	0.0244	0.0021	0.0285	0.0027	0.0330	0.0058
v_{11}	0.7545	0.0695	0.9332	0.1130	1.0811	0.1382
v_{12}	0.2648	0.0228	0.3059	0.0400	0.3511	0.0420
v_{13}	0.0934	0.0080	0.1048	0.0131	0.1244	0.0162
v_{14}	-0.4329	0.0343	-0.4402	0.0458	-0.5253	0.0630
v_{15}	0.3311	0.0271	0.3547	0.0366	0.4353	0.0558
v_{16}	0.5445	0.0488	0.5671	0.0547	0.7074	0.0558
v_{17}	0.5140	0.0531	0.5166	0.0544	0.6327	0.0490
v_{18}	0.1988	0.0104	0.2050	0.0154	0.2488	0.0279

Table B.13 Summary of average θ values for the 10 runs of test 14.

Input Variable	Test					
	14a		14b		14c	
	$t_f = 1\times10^{-5}$		$t_f = 1\times10^{-10}$		$t_f = 1\times10^{-20}$	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
v_1	26.1937	2.2401	24.8051	1.0237	25.6411	0.7016
v_2	49.5857	1.0105	49.1393	0.7083	49.4028	0.7036
v_3	61.0633	2.2543	60.9426	1.3084	60.8182	0.3325
v_4	-1.8E-08	1.08E-09	-1.5E-08	2.51E-09	-1.1E-08	5.51E-09
v_5	9.4E-09	1.31E-09	9.14E-09	1.88E-09	9.21E-09	1.95E-09
v_6	1.77E-09	1.68E-10	1.81E-09	2.64E-10	1.33E-09	3.15E-10
v_7	-4.8E-09	5.59E-10	-4.7E-09	1E-09	-3.9E-09	1.26E-09
v_8	-9.8E-09	9.71E-10	-1E-08	1.55E-09	-1.2E-08	1.14E-09
v_9	-9.4E-09	8.13E-10	-9.5E-09	1.3E-09	-8.8E-09	3.41E-09
v_{10}	-3.9E-09	6.57E-10	-4.2E-09	1.17E-09	-4E-09	9.24E-10
v_{11}	-1E-08	1.47E-09	-1E-08	1.7E-09	-1.3E-08	1.3E-09
v_{12}	1.24E-08	1.79E-09	1.2E-08	9.7E-10	7.85E-09	3.84E-09
v_{13}	-1.3E-08	8.18E-10	-1.3E-08	3.13E-09	-1.5E-08	2.1E-09
v_{14}	-3E-09	5.37E-10	-3E-09	5.72E-10	-2.9E-09	1E-09
v_{15}	2.96E-08	3.21E-09	3.26E-08	5.99E-09	3.72E-08	6.99E-09
v_{16}	10.3790	0.4862	10.2336	0.2733	10.2399	0.1671
v_{17}	12.9324	0.8689	12.7987	0.4738	12.2834	0.9426
v_{18}	43.0981	4.2001	41.0380	4.8602	42.3438	0.6670

Table B.14 Summary of average A values for the 10 runs of test 14.

Input Variable	Test					
	14a		14b		14c	
	$t_f = 1\times10^{-5}$		$t_f = 1\times10^{-10}$		$t_f = 1\times10^{-20}$	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
v_1	0.4118	0.0489	0.3642	0.0607	0.2888	0.0762
v_2	0.4181	0.0541	0.3631	0.0575	0.2788	0.0789
v_3	0.4392	0.0467	0.3903	0.0369	0.2942	0.0597
v_4	0.4349	0.0327	0.3557	0.0421	0.2314	0.0799
v_5	0.4315	0.0571	0.3622	0.0589	0.2230	0.1091
v_6	0.4450	0.0383	0.3725	0.0626	0.2312	0.0906
v_7	0.4129	0.0366	0.3794	0.0260	0.2061	0.0976
v_8	0.4458	0.0356	0.3898	0.0457	0.2851	0.0689
v_9	0.4281	0.0417	0.3594	0.0700	0.2498	0.0791
v_{10}	0.4295	0.0498	0.3598	0.0673	0.2411	0.1058
v_{11}	0.4198	0.0436	0.3775	0.0705	0.2614	0.0899
v_{12}	0.4365	0.0444	0.3390	0.0666	0.2976	0.0689
v_{13}	0.4143	0.0264	0.3701	0.0442	0.2678	0.0751
v_{14}	0.4200	0.0362	0.3616	0.0513	0.2666	0.0800
v_{15}	0.4354	0.0554	0.3838	0.0688	0.3094	0.0720
v_{16}	0.4167	0.0535	0.3384	0.0508	0.2820	0.0194
v_{17}	0.4292	0.0331	0.3672	0.0614	0.3148	0.0673
v_{18}	0.4381	0.0328	0.3662	0.0539	0.2815	0.0615

Table B.15 Summary of average B values for the 10 runs of test 14.

Input Variable	Test					
	14a		14b		14c	
	$t_f = 1\times10^{-5}$		$t_f = 1\times10^{-10}$		$t_f = 1\times10^{-20}$	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
v_1	0.1953	0.0033	0.1969	0.0024	0.1985	0.0020
v_2	0.1987	0.0018	0.1986	0.0027	0.1984	0.0021
v_3	0.1953	0.0054	0.1944	0.0081	0.1991	0.0009
v_4	0.1959	0.0048	0.1965	0.0063	0.1989	0.0011
v_5	0.1927	0.0075	0.1955	0.0052	0.1990	0.0009
v_6	0.1960	0.0041	0.1966	0.0036	0.1994	0.0011
v_7	0.1926	0.0069	0.1964	0.0035	0.1993	0.0010
v_8	0.1967	0.0058	0.1984	0.0032	0.1994	0.0012
v_9	0.1911	0.0095	0.1973	0.0018	0.1986	0.0010
v_{10}	0.1915	0.0128	0.1976	0.0033	0.1985	0.0023
v_{11}	0.1966	0.0041	0.1978	0.0036	0.1988	0.0023
v_{12}	0.1934	0.0073	0.1963	0.0042	0.1986	0.0021
v_{13}	0.1888	0.0098	0.1962	0.0030	0.1988	0.0010
v_{14}	0.1919	0.0100	0.1977	0.0039	0.1990	0.0010
v_{15}	0.1919	0.0085	0.1953	0.0088	0.1998	0.0004
v_{16}	0.1934	0.0064	0.1988	0.0018	0.1986	0.0012
v_{17}	0.1957	0.0045	0.1967	0.0026	0.1988	0.0015
v_{18}	0.1955	0.0053	0.1968	0.0054	0.1986	0.0022

Table B.16 Summary of average k values for the 10 runs of test 15.

Input Variable	Test					
	15a		15b		15c	
	$i_t = 20$		$i_t = 50$		$i_t = 100$	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
v_1	0.1761	0.0100	0.1797	0.0266	0.2102	0.0283
v_2	-0.7879	0.0511	-0.8704	0.1285	-0.9912	0.1715
v_3	0.6299	0.0425	0.5986	0.1117	0.6185	0.0986
v_4	-0.8318	0.0557	-0.8983	0.1173	-1.0247	0.1169
v_5	0.1076	0.0117	0.1297	0.0192	0.1474	0.0260
v_6	0.3676	0.0250	0.3995	0.0604	0.4410	0.0861
v_7	-0.0240	0.0018	-0.0226	0.0044	-0.0180	0.0048
v_8	0.9526	0.0760	0.9511	0.0821	1.0665	0.1666
v_9	-0.0045	0.0006	-0.0044	0.0007	-0.0049	0.0007
v_{10}	0.0244	0.0021	0.0264	0.0045	0.0279	0.0080
v_{11}	0.7545	0.0695	0.8553	0.0983	0.9272	0.1215
v_{12}	0.2648	0.0228	0.3125	0.0417	0.3556	0.0626
v_{13}	0.0934	0.0080	0.0948	0.012	0.1166	0.0162
v_{14}	-0.4329	0.0343	-0.4543	0.0490	-0.5146	0.0779
v_{15}	0.3311	0.0271	0.3913	0.0499	0.4322	0.0459
v_{16}	0.5445	0.0488	0.5531	0.0492	0.7126	0.1160
v_{17}	0.5140	0.0531	0.5475	0.0981	0.5869	0.1002
v_{18}	0.1988	0.0104	0.2141	0.0249	0.2266	0.0395

Table B.17 Summary of average θ values for the 10 runs of test 15.

Input Variable	Test					
	15a		15b		15c	
	$i_t = 20$		$i_t = 50$		$i_t = 100$	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
v_1	26.1937	2.2401	24.6377	2.1464	25.7843	3.0122
v_2	49.5857	1.0105	49.2757	1.4844	48.3011	1.7096
v_3	61.0633	2.2543	61.4230	1.1221	57.8878	10.2592
v_4	-1.8E-08	1.08E-09	-1.6E-08	2.29E-09	-1.4E-08	4.08E-09
v_5	9.4E-09	1.31E-09	8.57E-09	1.35E-09	8.22E-09	2.25E-09
v_6	1.77E-09	1.68E-10	1.89E-09	4.42E-10	1.53E-09	3.92E-10
v_7	-4.8E-09	5.59E-10	-4.5E-09	1.4E-09	-4.5E-09	1.96E-09
v_8	-9.8E-09	9.71E-10	-1E-08	1.83E-09	-1.3E-08	3.13E-09
v_9	-9.4E-09	8.13E-10	-9.9E-09	2.15E-09	-9.3E-09	2.22E-09
v_{10}	-3.9E-09	6.57E-10	-3.7E-09	5.98E-10	-4E-09	1.02E-09
v_{11}	-1E-08	1.47E-09	-1.1E-08	1.91E-09	-1.2E-08	2.49E-09
v_{12}	1.24E-08	1.79E-09	1.17E-08	1.95E-09	1.03E-08	2.69E-09
v_{13}	-1.3E-08	8.18E-10	-1.3E-08	2.4E-09	-1.5E-08	2.95E-09
v_{14}	-3E-09	5.37E-10	-2.6E-09	5.55E-10	-2.9E-09	8.27E-10
v_{15}	2.96E-08	3.21E-09	3.3E-08	5.53E-09	3.41E-08	1.17E-08
v_{16}	10.3790	0.4862	10.5987	0.5918	10.0726	0.6545
v_{17}	12.9324	0.8689	12.1744	1.2624	12.3357	0.4793
v_{18}	43.0981	4.2001	41.6448	3.2594	42.4410	1.1083

Table B.18 Summary of average A values for the 10 runs of test 15.

Input Variable	Test					
	15a		15b		15c	
	$i_t = 20$		$i_t = 50$		$i_t = 100$	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
v_1	0.4118	0.0489	0.3415	0.0701	0.3062	0.1072
v_2	0.4181	0.0541	0.3908	0.0405	0.2840	0.0993
v_3	0.4392	0.0467	0.3972	0.0442	0.3467	0.0877
v_4	0.4349	0.0327	0.3471	0.0579	0.2681	0.1209
v_5	0.4315	0.0571	0.3592	0.0536	0.2662	0.0858
v_6	0.4450	0.0383	0.3616	0.0869	0.2853	0.1058
v_7	0.4129	0.0366	0.3587	0.0676	0.3086	0.0980
v_8	0.4458	0.0356	0.3855	0.0807	0.2827	0.0803
v_9	0.4281	0.0417	0.3609	0.0852	0.2499	0.1399
v_{10}	0.4295	0.0498	0.3470	0.0764	0.2627	0.0670
v_{11}	0.4198	0.0436	0.4024	0.0428	0.2779	0.0992
v_{12}	0.4365	0.0444	0.3689	0.0482	0.2638	0.0702
v_{13}	0.4143	0.0264	0.3434	0.0796	0.2036	0.0962
v_{14}	0.4200	0.0362	0.3556	0.0564	0.2939	0.0778
v_{15}	0.4354	0.0554	0.3455	0.0789	0.2644	0.1136
v_{16}	0.4167	0.0535	0.3780	0.0709	0.2631	0.0799
v_{17}	0.4292	0.0331	0.3469	0.1025	0.2819	0.0862
v_{18}	0.4381	0.0328	0.3613	0.0799	0.3070	0.0877

Table B.19 Summary of average B values for the 10 runs of test 15.

Input Variable	Test					
	15a		15b		15c	
	$i_t = 20$		$i_t = 50$		$i_t = 100$	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
v_1	0.1953	0.0033	0.1944	0.0056	0.1971	0.0022
v_2	0.1987	0.0018	0.1949	0.0049	0.1971	0.0032
v_3	0.1953	0.0054	0.1928	0.0063	0.1938	0.0132
v_4	0.1959	0.0048	0.1959	0.0042	0.1962	0.0034
v_5	0.1927	0.0075	0.1901	0.0100	0.1893	0.0167
v_6	0.1960	0.0041	0.1896	0.0129	0.1833	0.0222
v_7	0.1926	0.0069	0.1911	0.0047	0.1910	0.0080
v_8	0.1967	0.0058	0.1973	0.0038	0.1967	0.0032
v_9	0.1911	0.0095	0.1905	0.0113	0.1859	0.0108
v_{10}	0.1915	0.0128	0.1885	0.0136	0.1900	0.0106
v_{11}	0.1966	0.0041	0.1931	0.0114	0.1975	0.0030
v_{12}	0.1934	0.0073	0.1926	0.0067	0.1929	0.0060
v_{13}	0.1888	0.0098	0.1883	0.0182	0.1891	0.0141
v_{14}	0.1919	0.0100	0.1931	0.0084	0.1906	0.0089
v_{15}	0.1919	0.0085	0.1950	0.0044	0.1960	0.0036
v_{16}	0.1934	0.0064	0.1943	0.0070	0.1924	0.0073
v_{17}	0.1957	0.0045	0.1919	0.0113	0.1973	0.0028
v_{18}	0.1955	0.0053	0.1950	0.0041	0.1913	0.0088

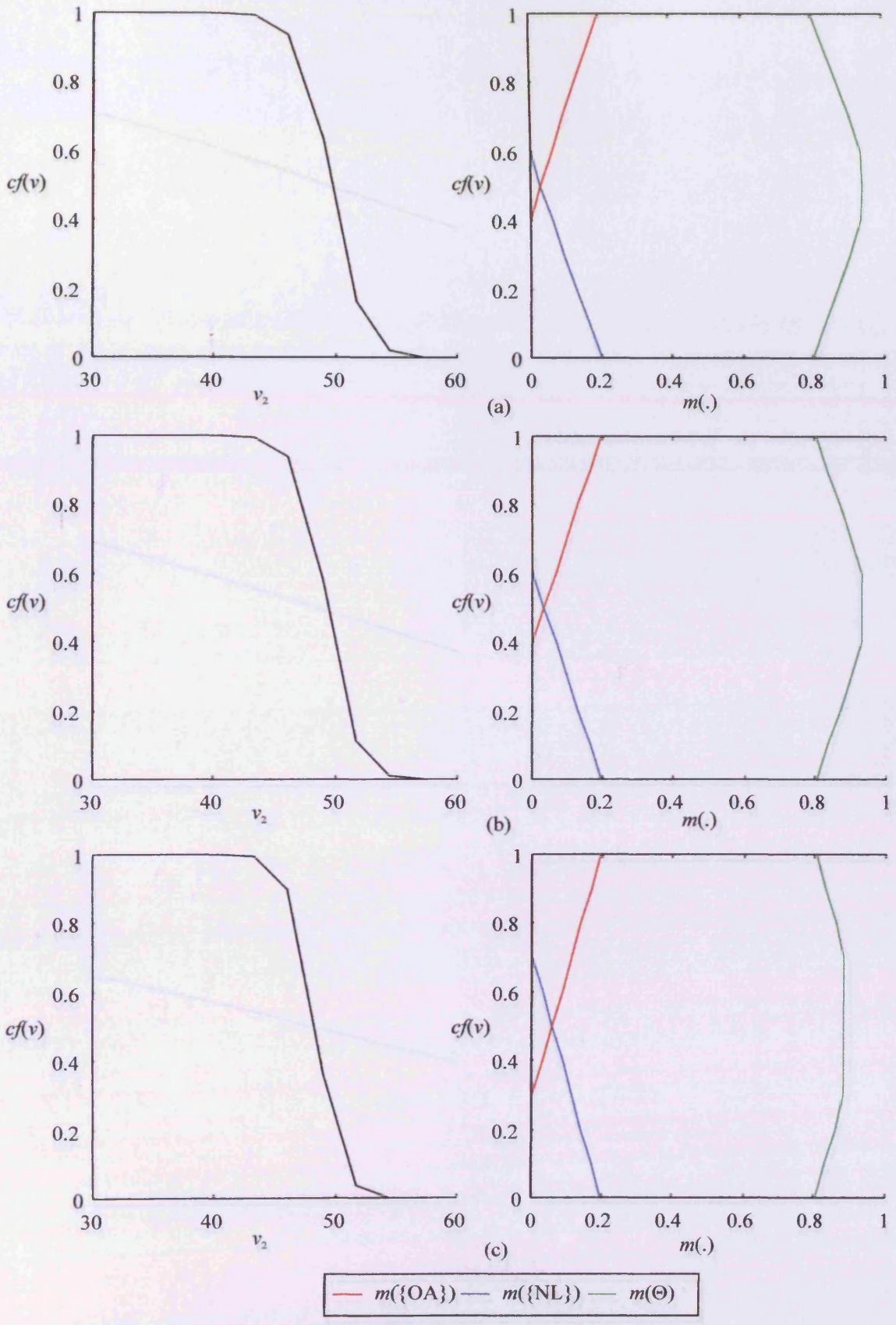


Figure B.97 The effect of decreasing the number of iterations per temperature, i_t on the confidence factor and BOE for v_2. (a) $i_t = 20$; (b) $i_t = 50$; (c) $i_t = 100$.

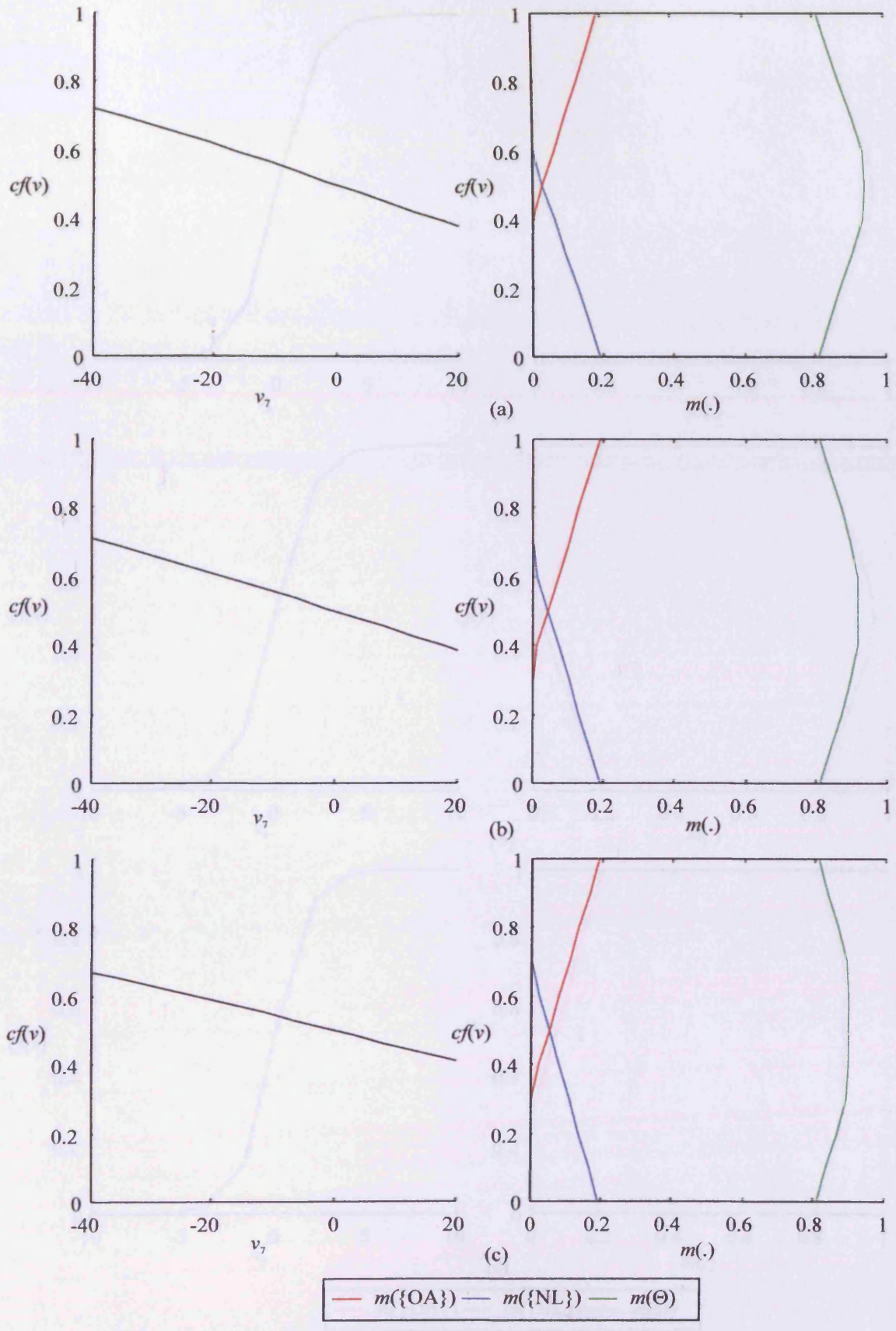


Figure B.98 The effect of decreasing the number of iterations per temperature, i_t on the confidence factor and BOE for v_7. (a) $i_t = 20$; (b) $i_t = 50$; (c) $i_t = 100$.

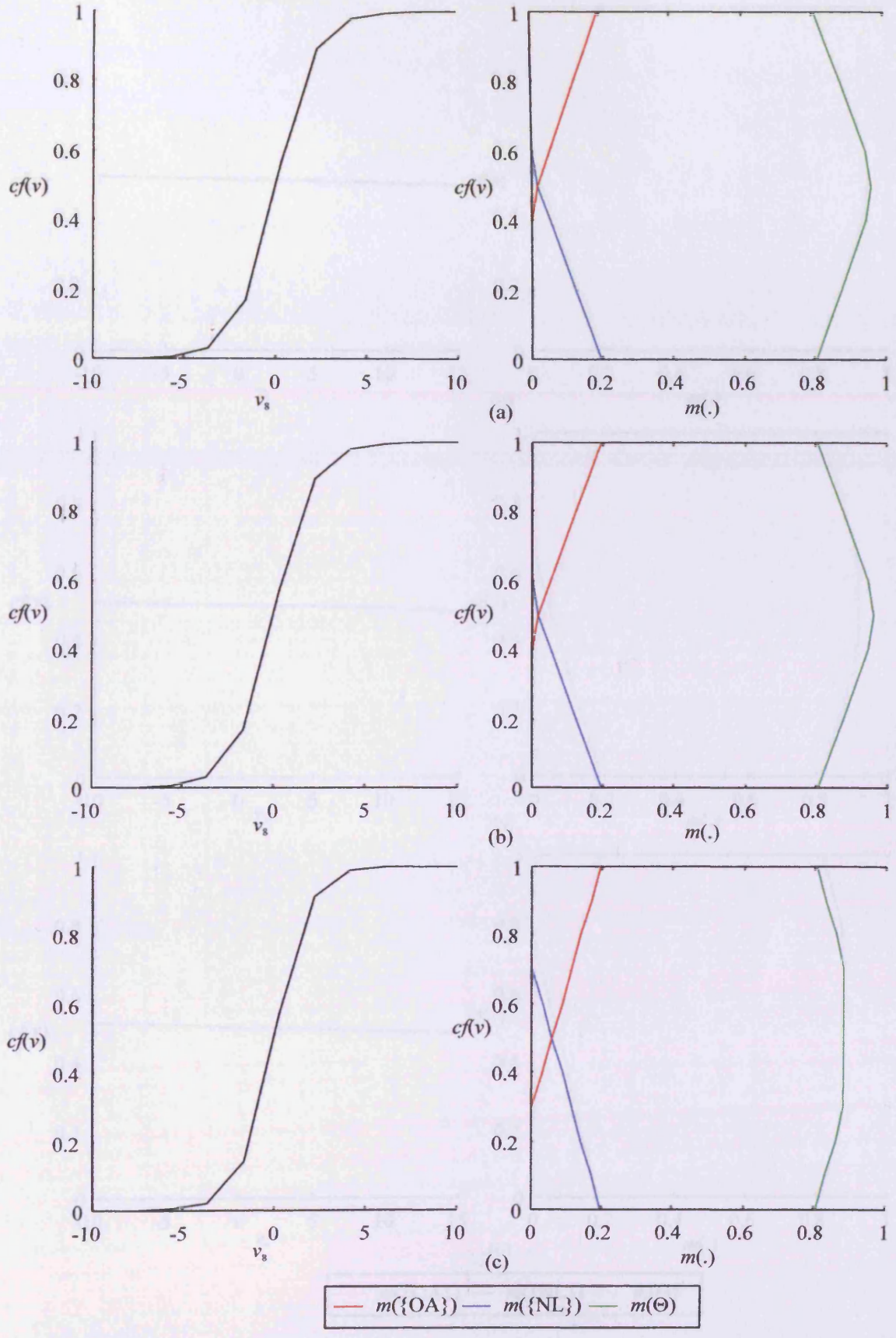


Figure B.99 The effect of decreasing the number of iterations per temperature, i_t on the confidence factor and BOE for v_8. (a) $i_t = 20$; (b) $i_t = 50$; (c) $i_t = 100$.

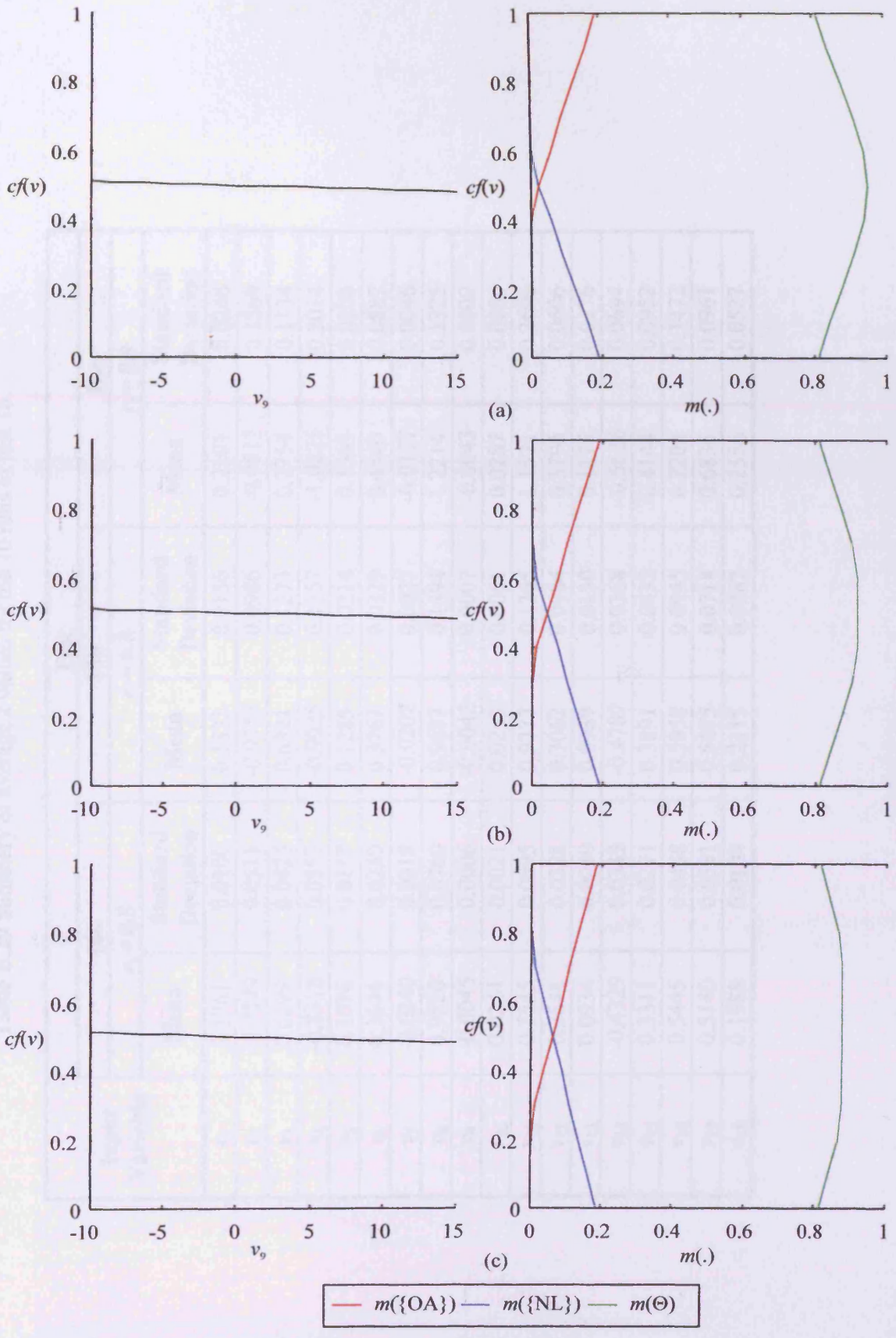


Figure B.100 The effect of decreasing the number of iterations per temperature, i_t on the confidence factor and BOE for v_9. (a) $i_t = 20$; (b) $i_t = 50$; (c) $i_t = 100$.

Table B.20 Summary of average k values for the 10 runs of test 16.

Input Variable	Test					
	16a		16b		16c	
	$r_t = 0.5$		$r_t = 0.8$		$r_t = 0.9$	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
v_1	0.1761	0.0100	0.1833	0.0156	0.2191	0.0246
v_2	-0.7879	0.0511	-0.9713	0.0986	-0.9512	0.1369
v_3	0.6299	0.0425	0.6321	0.1273	0.7754	0.1134
v_4	-0.8318	0.0557	-0.9025	0.1357	-1.0225	0.2014
v_5	0.1076	0.0117	0.1285	0.0214	0.1565	0.0208
v_6	0.3676	0.0250	0.3767	0.0329	0.4540	0.0882
v_7	-0.0240	0.0018	-0.0202	0.0027	-0.0177	0.0046
v_8	0.9526	0.0760	0.9687	0.1494	1.2214	0.1325
v_9	-0.0045	0.0006	-0.0042	0.0007	-0.0043	0.0009
v_{10}	0.0244	0.0021	0.0243	0.0051	0.0287	0.0080
v_{11}	0.7545	0.0695	0.9277	0.1365	1.1571	0.2696
v_{12}	0.2648	0.0228	0.3062	0.0484	0.3796	0.0606
v_{13}	0.0934	0.0080	0.0989	0.0130	0.1175	0.0176
v_{14}	-0.4329	0.0343	-0.4707	0.0368	-0.5626	0.0694
v_{15}	0.3311	0.0271	0.3891	0.0432	0.4144	0.0952
v_{16}	0.5445	0.0488	0.5958	0.0945	0.7203	0.1472
v_{17}	0.5140	0.0531	0.5475	0.0718	0.6835	0.0961
v_{18}	0.1988	0.0104	0.2115	0.0282	0.2556	0.0527

Table B.21 Summary of average θ values for the 10 runs of test 16.

Input Variable	Test					
	16a		16b		16c	
	$r_t = 0.5$		$r_t = 0.8$		$r_t = 0.9$	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
v_1	26.1937	2.2401	24.0912	2.8946	24.5994	1.7289
v_2	49.5857	1.0105	49.1173	0.5245	48.9006	0.4420
v_3	61.0633	2.2543	61.6312	1.6518	60.7485	0.2998
v_4	-1.8E-08	1.08E-09	-1.6E-08	3.86E-09	-1.3E-08	4.49E-09
v_5	9.4E-09	1.31E-09	8.97E-09	2.1E-09	1.01E-08	1.25E-09
v_6	1.77E-09	1.68E-10	1.9E-09	4.42E-10	1.48E-09	5.46E-10
v_7	-4.8E-09	5.59E-10	-4.8E-09	1.07E-09	-3.9E-09	2.04E-09
v_8	-9.8E-09	9.71E-10	-1E-08	1.88E-09	-1.3E-08	2.6E-09
v_9	-9.4E-09	8.13E-10	-8.5E-09	1.53E-09	-8.5E-09	2.24E-09
v_{10}	-3.9E-09	6.57E-10	-4.1E-09	6.82E-10	-3.9E-09	1.02E-09
v_{11}	-1E-08	1.47E-09	-1.1E-08	1.89E-09	-1.1E-08	2.55E-09
v_{12}	1.24E-08	1.79E-09	1.09E-08	2.28E-09	8.15E-09	4.11E-09
v_{13}	-1.3E-08	8.18E-10	-1.3E-08	2.33E-09	-1.6E-08	4.37E-09
v_{14}	-3E-09	5.37E-10	-2.8E-09	9.31E-10	-2.3E-09	6.54E-10
v_{15}	2.96E-08	3.21E-09	3.05E-08	8.44E-09	3.19E-08	9.54E-09
v_{16}	10.3790	0.4862	10.0085	0.8433	10.3653	0.4234
v_{17}	12.9324	0.8689	12.9133	0.7915	12.4773	0.1727
v_{18}	43.0981	4.2001	43.8823	6.4661	42.5013	0.6247

Table B.22 Summary of average A values for the 10 runs of test 16.

Input Variable	Test					
	16a		16b		16c	
	$r_t = 0.5$		$r_t = 0.8$		$r_t = 0.9$	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
v_1	0.4118	0.0489	0.3347	0.0826	0.2945	0.0866
v_2	0.4181	0.0541	0.3505	0.0660	0.2573	0.1258
v_3	0.4392	0.0467	0.3874	0.0778	0.2990	0.0707
v_4	0.4349	0.0327	0.3291	0.0905	0.2494	0.1040
v_5	0.4315	0.0571	0.3220	0.0850	0.2523	0.0805
v_6	0.4450	0.0383	0.3016	0.0756	0.2617	0.1245
v_7	0.4129	0.0366	0.3484	0.0726	0.2232	0.1256
v_8	0.4458	0.0356	0.3893	0.0550	0.2724	0.0742
v_9	0.4281	0.0417	0.3265	0.0601	0.1715	0.0962
v_{10}	0.4295	0.0498	0.3547	0.0667	0.2093	0.1269
v_{11}	0.4198	0.0436	0.3670	0.0523	0.2539	0.1113
v_{12}	0.4365	0.0444	0.3212	0.0616	0.2683	0.0472
v_{13}	0.4143	0.0264	0.3330	0.0933	0.1625	0.0931
v_{14}	0.4200	0.0362	0.3568	0.0548	0.2541	0.0786
v_{15}	0.4354	0.0554	0.3327	0.0668	0.2643	0.1073
v_{16}	0.4167	0.0535	0.3330	0.0729	0.2570	0.0346
v_{17}	0.4292	0.0331	0.3235	0.0733	0.3005	0.0445
v_{18}	0.4381	0.0328	0.3425	0.0712	0.2610	0.0624

Table B.23 Summary of average B values for the 10 runs of test 16.

Input Variable	Test					
	16a		16b		16c	
	$r_t = 0.5$		$r_t = 0.8$		$r_t = 0.9$	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
v_1	0.1953	0.0033	0.1961	0.0021	0.1956	0.0045
v_2	0.1987	0.0018	0.1949	0.0040	0.1988	0.0012
v_3	0.1953	0.0054	0.1882	0.0144	0.1973	0.0019
v_4	0.1959	0.0048	0.1950	0.0048	0.1980	0.0018
v_5	0.1927	0.0075	0.1862	0.0136	0.1867	0.0138
v_6	0.1960	0.0041	0.1944	0.0063	0.1941	0.0116
v_7	0.1926	0.0069	0.1892	0.0107	0.1932	0.0076
v_8	0.1967	0.0058	0.1929	0.0076	0.1982	0.0020
v_9	0.1911	0.0095	0.1935	0.0041	0.1961	0.0034
v_{10}	0.1915	0.0128	0.1892	0.0089	0.1934	0.0076
v_{11}	0.1966	0.0041	0.1974	0.0029	0.1964	0.0064
v_{12}	0.1934	0.0073	0.1923	0.0087	0.1931	0.0126
v_{13}	0.1888	0.0098	0.1922	0.0067	0.1944	0.0057
v_{14}	0.1919	0.0100	0.1965	0.0029	0.1967	0.0030
v_{15}	0.1919	0.0085	0.1932	0.0109	0.1964	0.0057
v_{16}	0.1934	0.0064	0.1924	0.0112	0.1955	0.0070
v_{17}	0.1957	0.0045	0.1937	0.0060	0.1911	0.0137
v_{18}	0.1955	0.0053	0.1940	0.0073	0.1927	0.0065

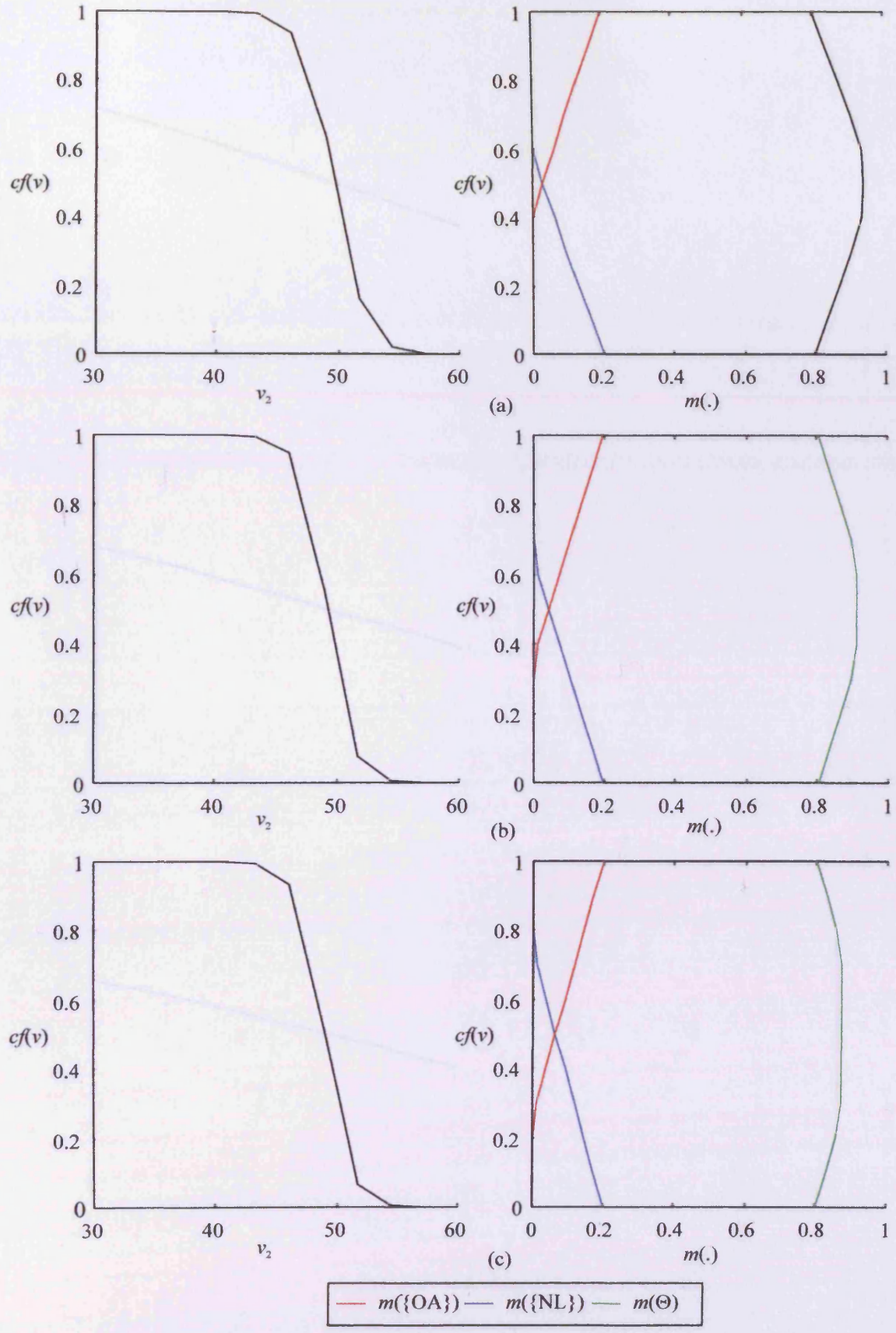


Figure B.101 The effect of decreasing the temperature reduction factor, r_t on the confidence factor and BOE for v_2. (a) $r_t = 0.5$; (b) $r_t = 0.8$; (c) $r_t = 0.9$.

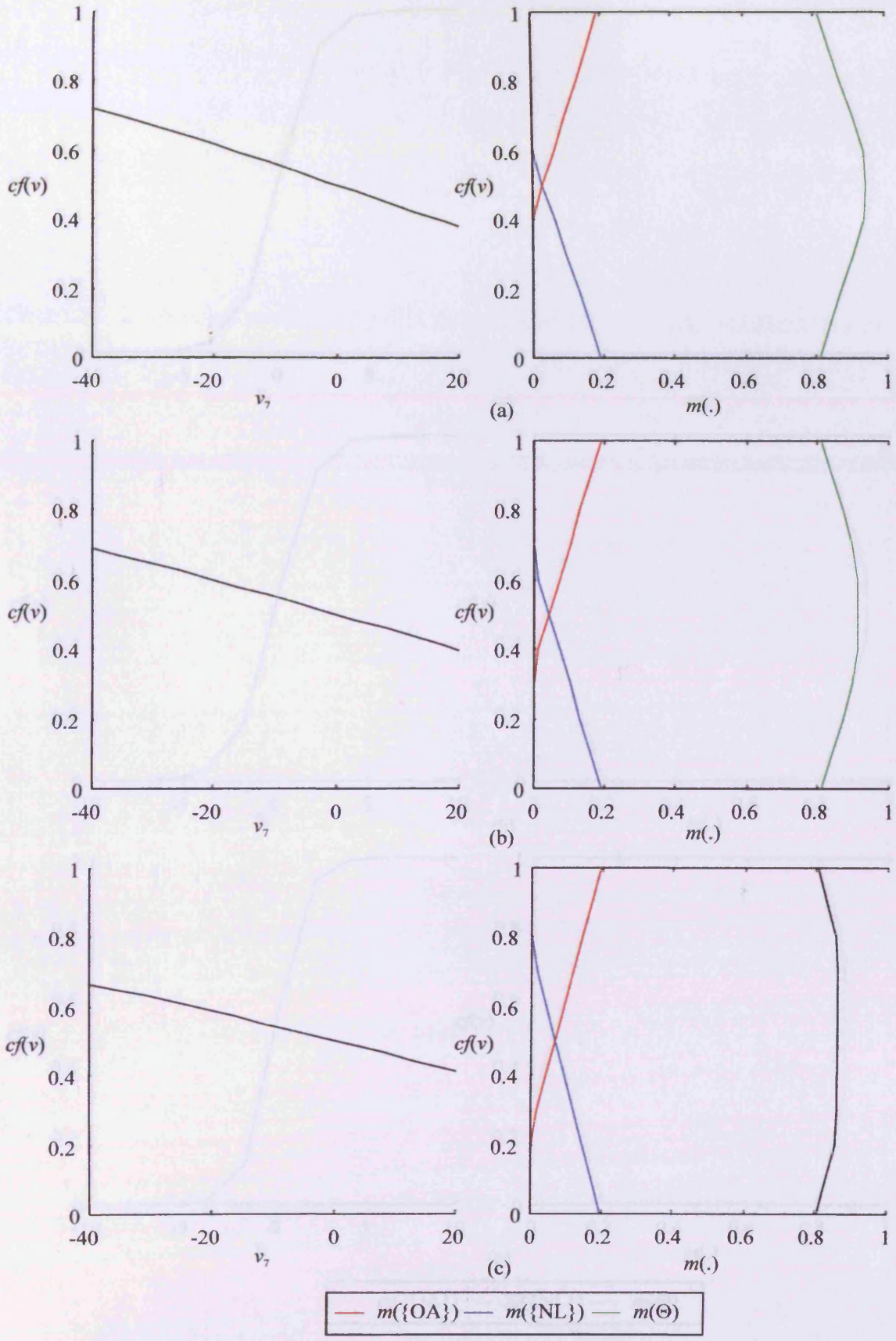


Figure B.102 The effect of decreasing the temperature reduction factor, r_t on the confidence factor and BOE for v_7. (a) $r_t = 0.5$; (b) $r_t = 0.8$; (c) $r_t = 0.9$.

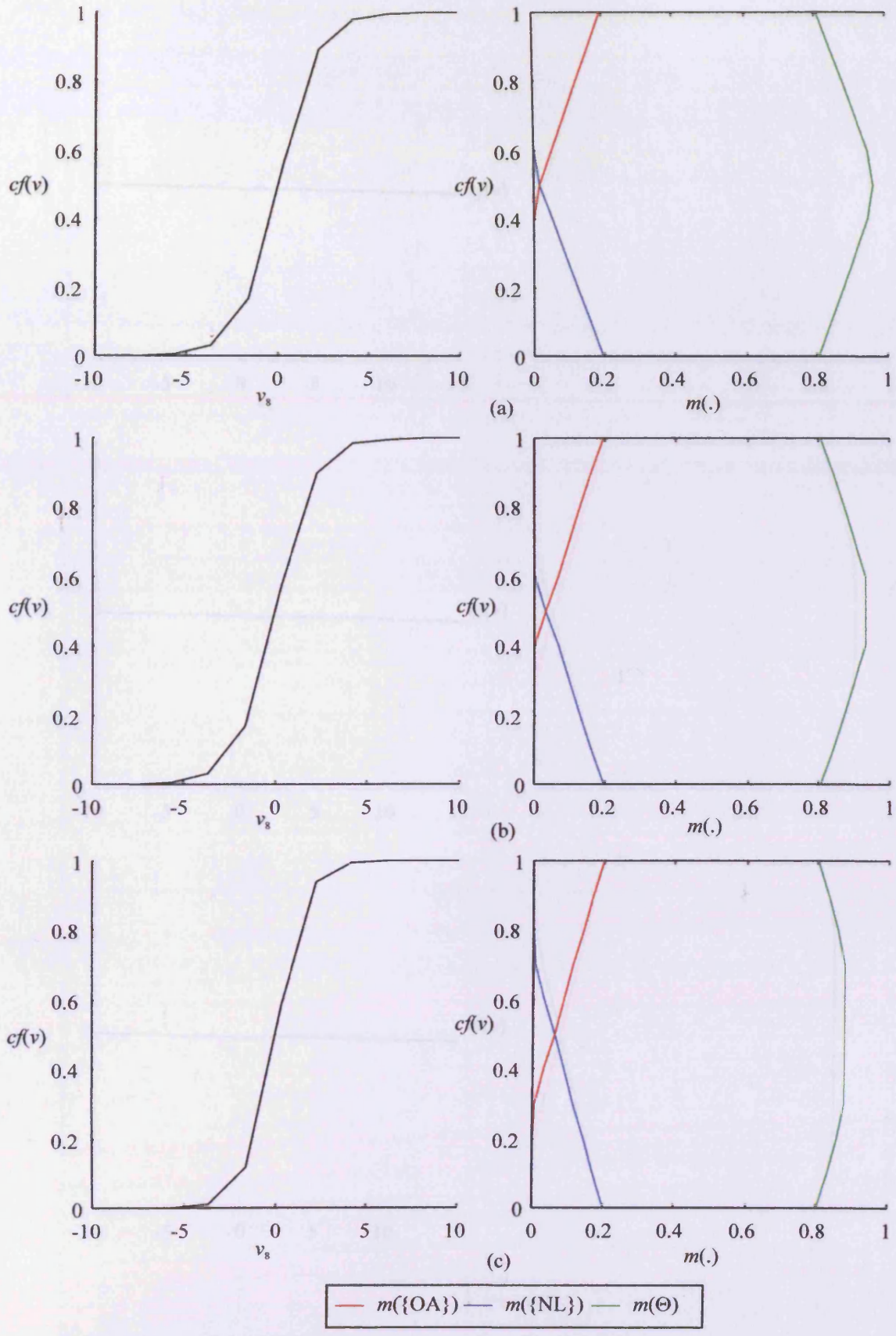


Figure B.103 The effect of decreasing the temperature reduction factor, r_t on the confidence factor and BOE for v_8. (a) $r_t = 0.5$; (b) $r_t = 0.8$; (c) $r_t = 0.9$.

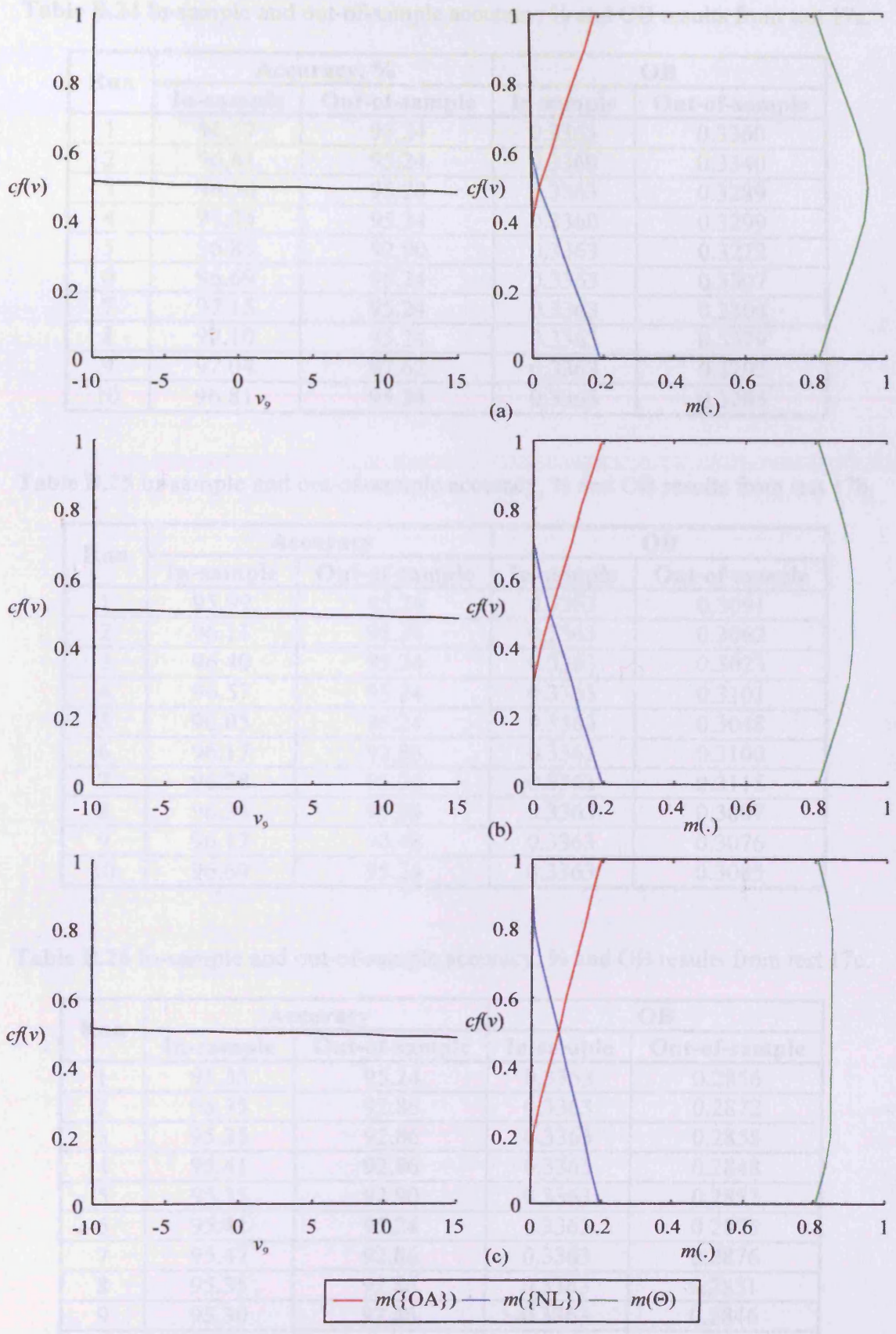


Figure B.104 The effect of decreasing the temperature reduction factor, r_t on the confidence factor and BOE for v_9. (a) $r_t = 0.5$; (b) $r_t = 0.8$; (c) $r_t = 0.9$.

Table B.24 In-sample and out-of-sample accuracy, % and OB results from test 17a.

Run	Accuracy, %		OB	
	In-sample	Out-of-sample	In-sample	Out-of-sample
1	96.57	95.24	0.3363	0.3360
2	96.81	95.24	0.3360	0.3340
3	96.50	95.20	0.3363	0.3289
4	97.04	95.24	0.3360	0.3299
5	96.86	92.90	0.3363	0.3272
6	96.69	95.24	0.3363	0.3307
7	97.15	95.24	0.3363	0.3304
8	97.10	95.24	0.3363	0.3379
9	97.04	97.62	0.3363	0.3292
10	96.81	95.24	0.3363	0.3285

Table B.25 In-sample and out-of-sample accuracy, % and OB results from test 17b.

Run	Accuracy		OB	
	In-sample	Out-of-sample	In-sample	Out-of-sample
1	95.99	95.24	0.3363	0.3091
2	96.11	95.24	0.3363	0.3062
3	96.40	95.24	0.3363	0.3023
4	96.57	95.24	0.3363	0.3101
5	96.05	95.24	0.3363	0.3048
6	96.17	92.86	0.3363	0.3100
7	96.28	95.24	0.3363	0.3115
8	96.34	95.24	0.3363	0.3007
9	96.17	90.48	0.3363	0.3076
10	96.69	95.24	0.3363	0.3085

Table B.26 In-sample and out-of-sample accuracy, % and OB results from test 17c.

Run	Accuracy		OB	
	In-sample	Out-of-sample	In-sample	Out-of-sample
1	95.35	95.24	0.3363	0.2856
2	95.35	92.86	0.3363	0.2872
3	95.35	92.86	0.3363	0.2855
4	95.41	92.86	0.3363	0.2848
5	95.35	92.90	0.3363	0.2853
6	95.41	95.24	0.3363	0.2875
7	95.47	92.86	0.3363	0.2876
8	95.35	92.86	0.3363	0.2831
9	95.30	92.86	0.3363	0.2846
10	95.24	92.86	0.3363	0.2863

Table B.27 In-sample and out-of-sample accuracy, % and OB results from test 18a.

Run	Accuracy		OB	
	In-sample	Out-of-sample	In-sample	Out-of-sample
1	96.57	95.24	0.3363	0.3360
2	96.81	95.24	0.3360	0.3340
3	96.50	95.20	0.3363	0.3289
4	97.04	95.24	0.3360	0.3299
5	96.86	92.90	0.3363	0.3272
6	96.69	95.24	0.3363	0.3307
7	97.15	95.24	0.3363	0.3304
8	97.10	95.24	0.3363	0.3379
9	97.04	97.62	0.3363	0.3292
10	96.81	95.24	0.3363	0.3285

Table B.28 In-sample and out-of-sample accuracy, % and OB results from test 18b.

Run	Accuracy		OB	
	In-sample	Out-of-sample	In-sample	Out-of-sample
1	96.40	95.24	0.3363	0.3086
2	96.52	95.24	0.3363	0.3123
3	96.17	95.24	0.3363	0.3163
4	96.28	97.62	0.3363	0.3147
5	96.52	95.24	0.3363	0.3107
6	96.46	95.24	0.3363	0.3127
7	96.57	95.24	0.3363	0.3115
8	96.40	95.24	0.3363	0.3128
9	96.28	95.24	0.3363	0.3161
10	96.40	95.24	0.3363	0.3126

Table B.29 In-sample and out-of-sample accuracy, % and OB results from test 18c.

Run	Accuracy		OB	
	In-sample	Out-of-sample	In-sample	Out-of-sample
1	95.76	92.86	0.3363	0.2884
2	95.82	92.86	0.3363	0.2936
3	95.53	92.86	0.3363	0.2985
4	95.76	92.86	0.3363	0.2932
5	95.76	95.24	0.3363	0.2925
6	95.88	90.48	0.3363	0.2947
7	95.99	97.62	0.3363	0.2960
8	95.70	92.86	0.3363	0.2966
9	95.82	95.24	0.3363	0.2977
10	95.53	95.24	0.3363	0.2906

Table B.30 In-sample and out-of-sample accuracy, % and OB results from test 19a.

Run	Accuracy		OB	
	In-sample	Out-of-sample	In-sample	Out-of-sample
1	96.57	95.24	0.3363	0.3360
2	96.81	95.24	0.3360	0.3340
3	96.50	95.20	0.3363	0.3289
4	97.04	95.24	0.3360	0.3299
5	96.86	92.90	0.3363	0.3272
6	96.69	95.24	0.3363	0.3307
7	97.15	95.24	0.3363	0.3304
8	97.10	95.24	0.3363	0.3379
9	97.04	97.62	0.3363	0.3292
10	96.81	95.24	0.3363	0.3285

Table B.31 In-sample and out-of-sample accuracy, % and OB results from test 19b.

Run	Accuracy		OB	
	In-sample	Out-of-sample	In-sample	Out-of-sample
1	96.23	95.24	0.3363	0.3035
2	96.46	95.24	0.3363	0.3101
3	96.57	95.24	0.3363	0.3087
4	96.57	92.86	0.3363	0.3134
5	96.05	92.86	0.3363	0.3131
6	96.11	95.24	0.3363	0.3029
7	96.05	95.24	0.3363	0.3065
8	96.11	95.24	0.3363	0.2994
9	96.28	95.24	0.3363	0.3094
10	96.23	92.86	0.3363	0.3147

Table B.32 In-sample and out-of-sample accuracy, % and OB results from test 19c.

Run	Accuracy		OB	
	In-sample	Out-of-sample	In-sample	Out-of-sample
1	95.59	95.24	0.3363	0.2860
2	95.24	92.86	0.3363	0.2885
3	95.35	95.24	0.3363	0.2873
4	95.41	95.24	0.3363	0.2830
5	95.47	95.24	0.3363	0.2877
6	95.47	92.86	0.3363	0.2882
7	95.53	92.86	0.3363	0.2881
8	95.64	95.24	0.3363	0.2882
9	95.76	95.24	0.3363	0.2893
10	95.47	92.86	0.3363	0.2850

APPENDIX C

Table C.1 In-sample and out-of-sample accuracy (%) results from test 1a (Training goal, MSE = 0.07)

Run	Accuracy, %	
	In-sample	Out-of-sample
1	74.10	54.76
2	74.22	50.00
3	73.11	47.62
4	74.39	54.76
5	74.62	45.24
6	73.69	54.76
7	73.40	54.76
8	74.27	52.38
9	74.10	73.81
10	74.04	52.38

Table C.2 In-sample and out-of-sample accuracy (%) results from test 1b (Training goal, MSE = 0.05)

Run	Accuracy, %	
	In-sample	Out-of-sample
1	76.07	61.90
2	75.67	52.38
3	75.38	54.76
4	75.20	54.76
5	75.20	54.76
6	74.85	64.29
7	75.32	54.76
8	75.03	52.38
9	76.13	57.14
10	75.61	52.38

Table C.3 In-sample and out-of-sample accuracy (%) results from test 1c (Training goal, MSE = 0.01)

Run	Accuracy, %	
	In-sample	**Out-of-sample**
1	77.93	73.81
2	77.64	61.90
3	77.93	54.76
4	77.82	73.81
5	77.93	61.90
6	77.82	69.05
7	77.70	55.52
8	77.64	61.90
9	77.87	66.67
10	77.87	59.52

Table C.4 In-sample and out-of-sample accuracy (%) results from test 1d (Training goal, MSE = 0.005)

Run	Accuracy, %	
	In-sample	**Out-of-sample**
1	77.93	61.90
2	77.93	59.53
3	77.93	52.38
4	77.93	61.90
5	77.93	50.00
6	77.93	61.90
7	77.93	61.90
8	77.93	59.52
9	77.93	52.38
10	77.93	61.90

Table C.5 In-sample and out-of-sample accuracy (%) results from test 1e (Training goal, MSE = 0.001)

Run	Accuracy, %	
	In-sample	**Out-of-sample**
1	77.93	61.90
2	77.93	57.14
3	77.93	57.14
4	77.93	64.29
5	77.93	59.52
6	77.93	64.29
7	77.93	59.52
8	77.93	50.00
9	77.93	64.29
10	77.93	57.14

Table C.6 In-sample and out-of-sample accuracy (%) results from test 2a (Learning rate, $\eta = 0.02$)

Run	Accuracy, %	
	In-sample	**Out-of-sample**
1	77.87	59.52
2	77.93	66.67
3	77.64	59.52
4	77.64	61.90
5	77.64	59.52
6	77.76	66.67
7	77.82	54.76
8	77.64	59.52
9	77.64	59.52
10	77.93	54.76

Table C.7 In-sample and out-of-sample accuracy (%) results from test 2b (Learning rate, $\eta = 0.1$)

Run	Accuracy	
	In-sample	**Out-of-sample**
1	77.93	73.81
2	77.64	61.90
3	77.93	54.76
4	77.82	73.81
5	77.93	61.90
6	77.82	69.05
7	77.70	55.52
8	77.64	61.90
9	77.87	66.67
10	77.87	59.52

Table C.8 In-sample and out-of-sample accuracy (%) results from test 2c (Learning rate, $\eta = 0.6$)

Run	Accuracy, %	
	In-sample	**Out-of-sample**
1	77.93	50.00
2	77.64	66.67
3	77.70	57.14
4	77.58	57.14
5	77.82	57.14
6	77.93	64.29
7	77.93	52.38
8	77.64	59.52
9	77.70	45.24
10	77.87	59.52

Table C.9 In-sample and out-of-sample accuracy (%) results from test 3a (Momentum constant, $\mu = 0.9$)

Run	Accuracy, %	
	In-sample	**Out-of-sample**
1	77.93	73.81
2	77.64	61.90
3	77.93	54.76
4	77.82	73.81
5	77.93	61.90
6	77.82	69.05
7	77.70	55.52
8	77.64	61.90
9	77.87	66.67
10	77.87	59.52

Table C.10 In-sample and out-of-sample accuracy (%) results from test 3b (Momentum constant, $\mu = 0.8$)

Run	Accuracy, %	
	In-sample	**Out-of-sample**
1	77.82	54.76
2	77.93	42.86
3	77.76	47.62
4	77.64	61.90
5	77.64	61.90
6	77.87	57.14
7	77.87	59.52
8	77.93	69.05
9	77.76	52.38
10	77.76	61.90

Table C.11 In-sample and out-of-sample accuracy (%) results from test 4a (Number of hidden neurons, $n_j = 5$)

Run	Accuracy, %	
	In-sample	**Out-of-sample**
1	77.93	73.81
2	77.64	61.90
3	77.93	54.76
4	77.82	73.81
5	77.93	61.90
6	77.82	69.05
7	77.70	55.52
8	77.64	61.90
9	77.87	66.67
10	77.87	59.52

Table C.12 In-sample and out-of-sample accuracy (%) results from test 4b (Number of hidden neurons, $n_j = 10$)

Run	Accuracy, %	
	In-sample	**Out-of-sample**
1	77.93	57.14
2	77.76	61.90
3	77.82	52.38
4	77.87	52.38
5	77.87	54.76
6	77.87	52.38
7	77.64	66.67
8	77.93	54.76
9	77.87	64.29
10	77.64	57.14

Table C.13 In-sample and out-of-sample accuracy (%) results from test 4c (Number of hidden neurons, $n_j = 15$)

Run	Accuracy, %	
	In-sample	Out-of-sample
1	77.87	64.29
2	77.82	54.76
3	77.87	52.38
4	77.87	52.38
5	77.87	52.38
6	77.93	54.76
7	77.93	64.29
8	77.93	57.14
9	77.93	61.90
10	77.93	52.38

Table C.14 In-sample and out-of-sample accuracy (%) results from test 4d (Number of hidden neurons, $n_j = 20$)

Run	Accuracy, %	
	In-sample	Out-of-sample
1	77.93	59.52
2	77.93	61.90
3	77.87	57.14
4	77.87	52.38
5	77.87	52.38
6	77.93	54.76
7	77.93	64.29
8	77.93	57.14
9	77.93	61.90
10	77.93	52.38

Table C.15 In-sample and out-of-sample accuracy (%) results from test 5a (Number of output neurons, $n_k = 1$)

Run	Accuracy, %	
	In-sample	Out-of-sample
1	77.93	73.81
2	77.64	61.90
3	77.93	54.76
4	77.82	73.81
5	77.93	61.90
6	77.82	69.05
7	77.70	55.52
8	77.64	61.90
9	77.87	66.67
10	77.87	59.52

Table C.16 In-sample and out-of-sample accuracy (%) results from test 5b (Number of output neurons, $n_k = 2$)

Run	Accuracy, %	
	In-sample	Out-of-sample
1	77.64	59.52
2	77.41	61.90
3	76.77	71.43
4	77.35	47.62
5	77.64	54.76
6	77.24	64.29
7	77.64	54.76
8	77.24	64.29
9	77.53	71.43
10	77.64	54.76